

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:39:16 ; Search time 32 Seconds

(without alignments)
715.038 Million cell updates/sec

Title: US-09-674-266a-181

Perfect score: 1063
Sequence: 1 RLSCAGTSLGSGPHSRRLT.....DTALLDNMKKALIKTEL 206

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.Geneseq_032802:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1063	100.0	206	20	AAV73837 Human prostate tum
2	1002	94.3	194	20	AAV75590 Human ovarian tumo
3	1002	94.3	202	22	AAV75621 Human colon cancer
4	899	84.6	175	19	AAW7365 zsig10 polypeptide
5	899	84.6	175	19	AAW37844 Human XAG growth f
6	899	84.6	175	19	AAW37872 Human protein comp
7	899	84.6	175	20	AAV59675 Secreted protein 1
8	899	84.6	175	21	AAAB24070 Human PRO1030 prot
9	899	84.6	175	22	AAU08804 Breast cancer cell
10	899	84.6	175	22	AAW72203 Human huxAG-1/CCSG
11	656	61.7	132	21	AAAG00109 Human secreted pro

12	605	56.9	131	20	AAV59718 Secreted protein 7
13	598	56.3	116	20	AAV12312 Human 5' EST secret
14	593	55.8	115	21	AAV64672 Human 5' EST relat
15	555	52.2	166	19	AAW37846 Human XAG growth f
16	555	52.2	166	21	AAAB00194 Breast cancer prot
17	555	52.2	166	22	AAU25728 Breast cancer-asso
18	555	52.2	166	22	AAW72205 Human huxAG-3/CCSG
19	555	52.2	166	22	AAAB31192 Amino acid sequenc
20	555	52.2	166	22	AAU07647 Human BCMP 11 poly
21	555	52.2	168	22	AAW24502 Human tumour relat
22	555	52.2	168	22	AAV76568 Human ovarian tumo
23	361	34.0	180	22	AAV73737 Human colon cancer
24	285	26.8	56	20	AAV11881 Human 5' EST secre
25	222	20.9	56	20	AAV11939 Human 5' EST secre
26	222	20.9	172	19	AAW37845 Human XAG growth f
27	222	20.9	172	20	AAV03236 Clone HP01766 of a
28	222	20.9	172	20	AAV03237 Clone HP01766 of a
29	222	20.9	172	20	AAV03230 Xenopus sp embryo
30	222	20.9	172	21	AAAB10282 Human PRO1376 prot
31	222	20.9	172	21	AAAB24332 Human membrane or
32	222	20.9	172	22	AAAB88366 Human huxAG-2/CCSG
33	222	20.9	172	22	AAW72204 Amino acid sequenc
34	218.5	20.6	172	19	AAW80812 Human 5' EST secre
35	210	19.8	62	20	AAV11882 Human 5' EST secre
36	179.5	16.9	70	20	AAV11654 Amino acid sequenc
37	88.5	8.3	118	20	AAV35816 Novel human diagno
38	88.5	8.3	1788	22	ABG06749 S cervistiae apopt
39	87.5	8.2	458	22	AAV070789 Novel human diagno
40	86.5	8.1	1614	22	ABG17640 Drosophila melano
41	84	7.9	2027	22	ABG07898 Novel human diagno
42	83.5	7.9	1409	22	ABAB65706 Novel human diagno
43	83	7.8	1146	22	ABG14251 Novel human diagno
44	82	7.7	556	22	ABG13300 Novel human diagno
45	82	7.7	595	22	ABG11877

ALIGNMENTS

RESULT 1
ID AAV73837 standard; Protein; 206 AA.
XX
AC AAV73837;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor EST fragment derived protein #24.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
XX
KW treatment.
XX
OS Homo sapiens.
XX
DE19820190-A1.
XX
PN
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (METRA-) METAGEN GBS GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-621386/54.
XX
DR N-PSDB; AA52865.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
XX
PS proteins -
XX
Claim 23; Page 318; 502pp; German.

XX This invention describes novel polypeptides and their encoding nucleic acids derived from human pancreatic tumor tissue which have cytostatic activity. The sequences are also useful in producing pharmaceutical compositions for treatment of pancreatic tumors. AAY73814-Y74252 CC represent protein fragments encoded by the human pancreatic tumor CDNA CC library derived expressed sequence tag (EST) sequences represented in CC AA52858-253014.

XX
SQ Sequence 206 AA;

Query Match 100.0%; Score 1063; DB 20; Length 206;
Best Local Similarity 100.0%; Pred. No. 2,3e-108;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSCAGTISGSGPHPSRLTQGRWVRSRVAMERIPVSAPLLVALSTLARDTTPKGA 60
DB 1 RLSCAGTISGSGPHPSRLTQGRWVRSRVAMERIPVSAPLLVALSTLARDTTPKGA 60

QY 61 KKDTRDSRPLPQTLRSRGWDQILMTQTYEALYKSKTSNKPMLIIHHLDECPSQALKK 120
DB 61 KKDTRDSRPLPQTLRSRGWDQILMTQTYEALYKSKTSNKPMLIIHHLDECPSQALKK 120

QY 121 VFAENKEIQKLAEOQVFLNLVYETTDKHLSPDQGYPRIMEVDSLVVRADITGRYSNRL 180
DB 121 VFAENKEIQKLAEOQVFLNLVYETTDKHLSPDQGYPRIMEVDSLVVRADITGRYSNRL 180

QY 181 YAEPADTALLDMKKALKLKTTEL 206
DB 181 YAEPADTALLDMKKALKLKTTEL 206

RESULT 2
AAY76590
ID AAY76590 standard; Protein: 194 AA.

XX AAY76590:
AC 10-Apr-2000 (first entry)
DT 10-Apr-2000 (first entry)
XX Human ovarian tumor EST fragment encoded protein 86.
DE Human ovarian tumor EST fragment encoded protein 86.
XX Expressed sequence tag; EST: human; ovarian tumor; anticancer;
KW gene therapy; treatment.
XX Homo sapiens.
XX DE19817557-A1.
XX 21-Oct-1999.
XX 09-Apr-1998; 98DE-1017557.
XX 09-Apr-1998; 98DE-1017557.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
PI WPI: 1999-591920/51.
XX N-PSDB: AA277486.
XX New nucleic acid sequences expressed in ovarian, and some other, cancer
PT tissues, and derived polypeptides, for treatment of ovarian cancer and
PT identification of therapeutic agents -
XX
PS Claim 25; Page 279; 310pp; German.

XX This invention describes novel nucleic acid (CDNA) sequences (A) which
CC have anticancer activity and are highly expressed in ovarian tumor
CC tissue (and some also in testis and breast cancer tissue). The products
CC of the invention can be used for gene therapy. (A) are used (1) for
CC recombinant expression of polypeptides (B) and (11) to isolate complete

CC genes. (B) are used (1) to identify agents suitable for treatment of
CC ovarian cancer; (11) directly for treating this form of cancer
CC (including expression from gene therapy vectors) and (11) for generation
CC of specific antibodies. (A) are identified by assembling ESTs (expressed
CC sequence tags) from a particular tissue type before comparison of
CC expression patterns. This allows a significantly longer fragment of the
CC gene to be revealed, so should reduce the number of failures associated
CC with the fact that ESTs from different libraries may represent different
CC parts of the same unknown gene, distorting the estimated frequency of
CC occurrence in a particular tissue. AAY76505-Y76638 represent protein
CC fragments encoded by the human ovarian tumor CDNA library derived EST
CC fragments represented in AA277450-277572.

XX
SQ Sequence 194 AA;

Query Match 94.3%; Score 1002; DB 20; Length 194;
Best Local Similarity 100.0%; Pred. No. 1e-101;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 PHSRRLTQGRWVRSRVAMERIPVSAPLLVALSTLARDTTPKGAKKDTRSPKLP 72
DB 1 PHSRRLTQGRWVRSRVAMERIPVSAPLLVALSTLARDTTPKGAKKDTRSPKLP 72

QY 73 QTLRSRGWDQILMTQTYEALYKSKTSNKPMLIIHHLDECPSQALKKVFENKEIQKLA 132
DB 73 QTLRSRGWDQILMTQTYEALYKSKTSNKPMLIIHHLDECPSQALKKVFENKEIQKLA 132

QY 61 QTLRSRGWDQILMTQTYEALYKSKTSNKPMLIIHHLDECPSQALKKVFENKEIQKLA 120
DB 61 QTLRSRGWDQILMTQTYEALYKSKTSNKPMLIIHHLDECPSQALKKVFENKEIQKLA 120

QY 133 EQVFLNLVYETTDKHLSPDQGYPRIMEVDSLVVRADITGRYSNRLYAEPADTALL 192
DB 133 EQVFLNLVYETTDKHLSPDQGYPRIMEVDSLVVRADITGRYSNRLYAEPADTALL 192

QY 121 EQVFLNLVYETTDKHLSPDQGYPRIMEVDSLVVRADITGRYSNRLYAEPADTALL 180
DB 121 EQVFLNLVYETTDKHLSPDQGYPRIMEVDSLVVRADITGRYSNRLYAEPADTALL 180

QY 193 DNKKALKLKTTEL 206
DB 193 DNKKALKLKTTEL 206

RESULT 3
AAG75621
ID AAG75621 standard; Protein: 202 AA.

XX AAG75621:
AC 03-SEP-2001 (first entry)
DT 03-SEP-2001 (first entry)
XX Human colon cancer antigen protein SEQ ID NO:6385.
DE Human colon cancer antigen protein SEQ ID NO:6385.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-Apr-2001.
XX 28-SEP-2000; 2000WO-US26524.
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI: 2001-235357/24.
XX N-PSDB: AAH35026.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 7865-7866; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAM7196 to AAM7204
 CC and AAM7789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 202 AA:

Query Match 94.3%; Score 1002; DB 22; Length 202;

Best Local Similarity 100.0%; Pred. No. 1.1e-101;

Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 PPHSRRLTQGRVNRKSRVAMEKIPVSAPFLLVASTYLARDTTVKGAKKDKDSRPKLP 72
 DB 9 PPHSRRLTQGRVNRKSRVAMEKIPVSAPFLLVASTYLARDTTVKGAKKDKDSRPKLP 68
 OY 73 QTLRSGWGDQILMTQTYEALYKSKTSNKPMLIHHLEDCPHSQALKKVFANKEIOKLA 132
 DB 69 QTLRSGWGDQILMTQTYEALYKSKTSNKPMLIHHLEDCPHSQALKKVFANKEIOKLA 128
 OY 133 EGFVLNLYVETTDKHLSPDGOYVPRIMFVDPSTLVVRADITGRYSNRLAYEPADTALL 192
 DB 129 EGFVLNLYVETTDKHLSPDGOYVPRIMFVDPSTLVVRADITGRYSNRLAYEPADTALL 188
 OY 193 DNMRKALKLKTTEL 206
 DB 189 DNMRKALKLKTTEL 202

RESULT 4

AAM77365
 ID AAM77365 standard; Protein; 175 AA.

AC AAM77365;

DT 21-DEC-1998 (first entry)

XX zsig10 polypeptide.

XX Human; mucous-mediated function; adhesion; tumour metastasis;
 KW bacterial colonisation; microbial infection; AIDS; cystic fibrosis;
 KW chronic obstructive pulmonary disease; asthma; Crohn's disease;
 KW sinonasal inflammatory disease; inflammatory bowel disease; bronchitis.

XX Homo sapiens.

PN MO9841627-A1.

PD 24-SEP-1998.

PE 18-MAR-1998; 98WO-US05251.

PR 19-MAR-1997; 97US-0039631.

PA (ZYMO) ZYMOGENETICS INC.

PI Sheppard PO;

XX WPI; 1998-531566/45.

DR N-PSDB; AAV59320.

XX New isolated mucous-associated polypeptide, zsig10 - used to develop
 PT products for treating e.g. tumour metastasis, microbial infections,
 PT cystic fibrosis, asthma, bronchitis or inflammatory bowel disease
 XX
 XX Claim 1; Page 82; 109pp; English.

CC The human polypeptide zsig10 is involved in mucous-mediated functions
 CC such as adhesion. The products of the invention can be used in the study
 CC and treatment of e.g. tumour metastasis, bacterial colonisation,
 CC susceptibility to and persistence of infection, microbial infections,
 CC AIDS, cystic fibrosis, chronic obstructive pulmonary disease, asthma,
 CC sinonasal inflammatory disease, inflammatory bowel disease, bronchitis,
 CC or Crohn's disease. The products can also be used for detection,
 CC diagnosis and drug screening.

Sequence 175 AA:

Query Match 84.6%; Score 899; DB 19; Length 175;

Best Local Similarity 100.0%; Pred. No. 1.8e-90;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 MEKIPVSAPFLLVASTYLARDTTVKGAKKDKDSRPKLPQTLRSGWGDQILMTQTYEE 91
 DB 1 MEKIPVSAPFLLVASTYLARDTTVKGAKKDKDSRPKLPQTLRSGWGDQILMTQTYEE 60
 OY 92 ALYKSKTSNKPMLIHHLEDCPHSQALKKVFANKEIOKLAQFVLNLYVETTDKHLSP 151
 DB 61 ALYKSKTSNKPMLIHHLEDCPHSQALKKVFANKEIOKLAQFVLNLYVETTDKHLSP 120
 OY 152 DGOYVPRIMFVDPSTLVVRADITGRYSNRLAYEPADTALLDNMRKALKLKTTEL 206
 DB 121 DGOYVPRIMFVDPSTLVVRADITGRYSNRLAYEPADTALLDNMRKALKLKTTEL 175

RESULT 5

AAM37844
 ID AAM37844 standard; Protein; 175 AA.

AC AAM37844;

DT 28-AUG-1998 (first entry)

XX Human XAG growth factor huxAG-1.

XX huxAG-1; XAG; growth factor; colon cancer; tumour marker;

KW breast disease; liver disease; lung disease; emphysema;

KW wound healing; diagnosis; therapy; human.

XX Homo sapiens.

PH Key Location/Qualifiers

FT Peptide 1..20 /label= Sig_peptide

FT Protein 21..175 /label= Mat.protein

FT Peptide /note= "Claim 11"

FT Peptide /label= Epitope 20..43

FT Peptide /note= "Claim 12" 44..52

FT Peptide /label= Epitope 61..72

FT Peptide /note= "Claim 12" 90..103

FT Peptide /label= Epitope /note= "Claim 12" 113..125

FT Peptide /note= "Claim 12" 138..150

```

FT  /label= Epitope
FT  /note= "Claim 12"
XX
PN  MO9807749-A1.
XX
PD  26-FEB-1998.
XX
PF  22-AUG-1997; 97WO-US14139.
XX
PR  23-AUG-1996; 96WO-US13766.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Dillion PJ, Ebner R, Endress GA, Yu G;
XX  WPI: 1998-169093/15.
DR  N-PSDB: AAV19155.
XX
PT  New isolated human XAG growth factor(s) - used to develop products
PT  for treating e.g. liver, lung or breast diseases or
PT  hyperproliferative disorders, e.g. cancer.
XX
PS  Claim 1; Fig 1; 141pp; English.
XX
CC  This polypeptide comprises huxAG-1, a member of a novel family of
CC  human growth factors also including huxAG-2 (see AAW37845) and
CC  huxAG-3 (see AAW37846). These proteins share homology with the
CC  XAG protein of Xenopus laevis, which is involved in embryogenesis
CC  and is expressed in adult tissue. huxAG-1 is specifically found
CC  in cancerous colon cells and may therefore be a growth factor for
CC  colon cancer. huxAG-1 cDNA (see AAV19155) was isolated from a cDNA
CC  library derived from human colon cancer tissue. Vectors, host
CC  cells, antibodies, and screening methods for identifying agonists
CC  and antagonists of huxAG-1 are provided. HuxAG polypeptides are
CC  growth factors and can be used to stimulate proliferation of cells.
CC  They can be used to stimulate the proliferation and differentiation
CC  of hepatocytes to alleviate or treat liver diseases and pathologies
CC  such as fulminant liver failure caused by cirrhosis, liver damage
CC  caused by viral hepatitis and toxic substances. They can also be
CC  used to stimulate or promote liver regeneration, e.g. after
CC  surgery. They can also be used to prevent and heal damage to the
CC  lungs caused by various pathological states. They can be used to
CC  stimulate proliferation and differentiation and promote the repair
CC  of alveoli and bronchiolar epithelium to prevent, attenuate, or
CC  treat acute or chronic lung damage, e.g. emphysema, which results
CC  in the progressive loss of alveoli, and inhalation injuries, e.g.
CC  resulting from smoke inhalation and burns, that cause necrosis of
CC  the bronchiolar epithelium and alveoli. They can also be used to
CC  stimulate the proliferation and differentiation of breast tissue
CC  and could therefore be used to promote healing of breast tissue
CC  injury due to surgery, trauma or cancer. Antagonists can be used
CC  to treat hyperproliferative disorders, including cancer, in
CC  particular hepatocellular carcinoma, osteosarcoma, breast cancer,
CC  or colon cancer. The products can also be used for detection and
CC  diagnosis.
XX
SQ  Sequence 175 AA;

Query Match 84.6%; Score 899; DB 19; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.8e-90;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 6
AAW37872
ID  AAW37872 standard; Protein: 175 AA.
XX
AC  AAW37872;
XX
DT  10-AUG-1998 (first entry)
XX
DE  Human protein comprising secretory signal amino acid sequence 9.
XX
KW  Human protein; secretory signal; nutritional source; cytokine;
KW  immunity; haematopoiesis; activin; inhibin; tumour; chemotactic;
KW  chemokine; thrombolytic; anti-inflammatory; inhibition;
KW  stomach cancer cell.
XX
OS  Homo sapiens.
XX
XX
XX  WO9811217-A2.
XX
PD  19-MAR-1998.
XX
XX  12-SEP-1997; 97WO-JP03239.
XX
PR  13-SEP-1996; 96JP-0243060.
XX
PA  (PROT-) PROMEGENE INC.
PA  (SKGA ) SAGAMI CHEM RES CENTRE.
XX
PI  Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX  WPI: 1998-207380/18.
DR  N-PSDB: AAV29047, AAV29048.
XX
PT  Human proteins with secretory signal sequences - used to treat
PT  immune deficiencies, infections, tumours, and haematopoietic
PT  disorders, etc.
XX
PS  Claim 1; Pages 79; 131pp; English.
XX
CC  This is the amino acid sequence of a novel human protein comprising
CC  a secretory signal isolated from stomach cancer cells. Its proteins
CC  can be used as nutritional sources or supplements. The proteins may
CC  also have cytokine functions, immune modulating functions,
CC  haematopoiesis regulating activity, activin/inhibin regulating
CC  activity, chemotactic/chemokine activity, haemostatic and
CC  thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC  activity, tumour inhibition activity.
XX
SQ  Sequence 175 AA;

Query Match 84.6%; Score 899; DB 19; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.8e-90;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



```

XX AC AAY59675;
XX DT 18-JAN-2000 (first entry)
XX DE Secreted protein 108-008-5-0-A6-FL.
XX KW Secreted protein; fingerprint identification technique;
XX KW chromosome mapping; human; hereditary disease; diagnosis; cancer;
XX KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
XX KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
XX KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
XX KW hypertension.
XX OS Homo sapiens.
XX PN W09940189-A2.
XX PD 12-AUG-1999.
XX PF 09-FEB-1999; 99WO-1B00282.
XX PR 09-FEB-1998; 98US-0074121.
XX PR 13-APR-1998; 98US-0081563.
XX PR 10-AUG-1998; 98US-0096116.
XX PR 04-SEP-1998; 98US-0099273.
XX PA (GIST ) GENSET.
XX PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX DR MPI: 1999-600966/51.
XX DR N-PSDB; AAZ40803.
XX PT Extended cDNAs useful for expressing secreted proteins and to obtain
XX PT specific antibodies -
XX PS Claim 10; Page 199; 244pp; English.
XX CC This sequence represents a human secreted protein of the invention.
XX CC The extended cDNAs (or genomic DNAs obtainable from them) may be used to
XX CC prepare PCR primers and probes. These are useful for forensic matching or
XX CC positive identification by DNA sequencing. They may also be used in
XX CC alternative fingerprint identification techniques. Antibodies against the
XX CC proteins encoded by the extended cDNAs are useful in identification of
XX CC tissue types or cell species, as well as identifying tissue specific
XX CC soluble proteins. The sequences can be used for chromosome mapping and
XX CC identification of genes associated with hereditary diseases or drug
XX CC response. Signal sequences from the cDNAs can be used in construction of
XX CC secretion vectors. Other sequences derived from the extended cDNAs can be
XX CC used to clone upstream genomic DNA sequences including promoters. This is
XX CC in turn useful for identifying proteins that interact with promoter
XX CC sequences. Some of the proteins may be useful in diagnosing and treating
XX CC several disorders including, but not limited to: cancer, hyperlipidaemia,
XX CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
XX CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
XX CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX SQ Sequence 175 AA;

```

```

Query Match 84.6%; Score 899; DB 20; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.8e-90;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 32 MEKIPVSAFLLLVALSYTLARDTVKPGAKKDTKDSRPKLPOTLSRGWGDOLIMTQTYEE 91
DB 1 MEKIPVSAFLLLVALSYTLARDTVKPGAKKDTKDSRPKLPOTLSRGWGDOLIMTQTYEE 60
OY 92 ALYKSTSKRPMLTIHHHDECPHSQALKKVPFAENKRIQKLAQFVLNLVYETTDKHLSP 151
DB 61 ALYKSTSKRPMLTIHHHDECPHSQALKKVPFAENKRIQKLAQFVLNLVYETTDKHLSP 120
OY 152 DQGYVRIMFVDPSTLVRADITGRYSNRLYAVEPADTALLDMKKALKLKTLEL 206

```

```

DB 121 DQGYVRIMFVDPSTLVRADITGRYSNRLYAVEPADTALLDMKKALKLKTLEL 175
RESULT 8
AAB24070
ID AAB24070 standard; Protein; 175 AA.
XX AC AAB24070;
XX DT 29-JAN-2001 (first entry)
XX DE Human PRO1030 protein sequence SEQ ID NO:40.
XX KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
XX KW proliferation; tumorigenesis; identification; cancer; cytostatic;
XX KW nocrotropic; neuroprotective; antiinflammatory; immunosuppressive;
XX KW immunostimulant; antitumorigenic; leukaemia; lymphoid malignancy;
XX KW neuronal disorder; glial disorder; astrocytic disorder; angiogenic;
XX KW hypothalamic disorder; glandular disorder; macrophagal disorder;
XX KW epithelial disorder; stromal disorder; blastocoelec disorder;
XX KW inflammatory disorder; immunologic disorder.
XX OS Homo sapiens.
XX PN W0200053755-A2.
XX PD 14-SEP-2000.
XX PF 06-JAN-2000; 2000WO-US00376.
XX PR 08-MAR-1999; 99WO-US05028.
XX PR 02-JUN-1999; 99WO-US12252.
XX PR 23-JUN-1999; 99US-0141037.
XX PR 07-JUL-1999; 99US-0143048.
XX PR 26-JUL-1999; 99US-0145698.
XX PR 30-NOV-1999; 99WO-US28313.
XX PR 20-DEC-1999; 99WO-US30911.
XX PR 05-JAN-2000; 2000WO-US00219.
XX PA (GENTH ) GENENTECH INC.
XX PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillian KJ, Roy MA;
XX PI Watanabe CK, Wood WI;
XX DR MPI: 2000-572270/53.
XX DR N-PSDB; AAC58380.
XX PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
XX PT treatment, diagnosis and prevention of cancer -
XX PS Claim 61; Fig 28; 286pp; English.
XX CC The present invention describes an isolated antibody that binds to
XX CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
XX CC PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
XX CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1133, PRO1182, PRO1184,
XX CC PRO1187, PRO1281, PRO1283, PRO1284, PRO1317, PRO1710, PRO2094,
XX CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
XX CC growth. The PRO polypeptides and nucleotides are useful in the
XX CC treatment, diagnosis and prevention of cancer. The antibodies and other
XX CC anti-tumour compounds may be used to treat various conditions, including
XX CC those characterised by overexpression and/or activation of the amplified
XX CC PRO genes. Exemplary conditions or disorders to be treated with such
XX CC antibodies and other compounds include benign or malignant tumours
XX CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
XX CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
XX CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
XX CC leukaemias, and lymphoid malignancies, other disorders such as neuronal,
XX CC glial, astrocytic, hypothalamic and other glandular, macrophagal,
XX CC epithelial, stromal and blastocoelec disorders, and inflammatory,
XX CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
XX CC primers and hybridisation probes used in the isolation of the human PRO

```

CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.

XX Sequence 175 AA:

Query Match 84.6%; Score 899; DB 21; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.8e-90;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPYSAFLILVALSYTLARDTTVRGAKKDKDSRPKLPQTLRSGWGQDLIMTQYEE 91
DB 1 MEKIPYSAFLILVALSYTLARDTTVRGAKKDKDSRPKLPQTLRSGWGQDLIMTQYEE 60
QY 92 ALYKSTSNKPLMIHHLDECPSQALKVFAENKEIQKLAEOFVLNLVYETTDKHLSP 151
DB 61 ALYKSTSNKPLMIHHLDECPSQALKVFAENKEIQKLAEOFVLNLVYETTDKHLSP 120
QY 152 DGOYPRIMFVDPSTLYRADITGRYSNRLYAYEPADTALLDMKKALKLKTTEL 206
DB 121 DGOYPRIMFVDPSTLYRADITGRYSNRLYAYEPADTALLDMKKALKLKTTEL 175

RESULT 9

AAU08804
ID AAU08804 standard; Protein: 175 AA.

XX AC AAU08804;

DT 19-DEC-2001 (first entry)

XX DE Breast cancer cell membrane associated protein 7 (BCMP 7).

XX KM Breast cancer cell membrane protein 7; BCMP 7; breast cancer;
XX metastasis; cytostatic; gene therapy; antibody; antisense;
XX chromosome 7p21.3; vaccine.

XX OS Homo sapiens.

XX FT Key Location/Qualifiers

XX FT Peptide 1..20

XX FT Protein /note- "Signal peptide"

XX FT Protein /note- "Mature BCMP 7"

XX PN WC200163290-A1.

XX PD 30-AUG-2001.

XX PF 21-FEB-2001; 2001MO-GB00734.

XX PR 25-FEB-2000; 2000GB-0004576.

XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX PI Boyd RS, Stamps AC, Terrett JA, Tyson KL;

XX DR WPI; 2001-570651/64.

XX DR N-PSDB; AAS13480.

PT Diagnosing, preventing and treating breast cancer using a breast cancer
PT cell membrane protein BCMP 7 -

XX PS Claim 1; Fig 1; 62pp; English.

CC The invention describes the novel use of a protein found in breast
CC cancer cell membranes (BCMP 7) for diagnosing, preventing and treating
CC breast cancers. The peptide has cytostatic action and potential uses in
CC gene therapy and in vaccines. The polypeptide, antisense nucleic acids,
CC or fusion proteins comprising and Green fluorescent protein or the DsRed
CC fluorescent protein, antibodies specific for and/or nucleic acid are
CC used for the prevention and/or treatment of breast cancer. Antibodies
CC against may also be used for screening for and/or diagnosis of breast

CC cancer in a patient. The method for monitoring/assessing breast cancer
CC treatment in a patient and for the identification of metastatic breast
CC cancer cells in samples from a patient. This sequence is breast
CC cancer cell associated protein 7 (BCMP 7), encoded by a gene located on
CC chromosome 7p21.3, described in the method of the invention.

XX Sequence 175 AA:

Query Match 84.6%; Score 899; DB 22; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.8e-90;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPYSAFLILVALSYTLARDTTVRGAKKDKDSRPKLPQTLRSGWGQDLIMTQYEE 91
DB 1 MEKIPYSAFLILVALSYTLARDTTVRGAKKDKDSRPKLPQTLRSGWGQDLIMTQYEE 60
QY 92 ALYKSTSNKPLMIHHLDECPSQALKVFAENKEIQKLAEOFVLNLVYETTDKHLSP 151
DB 61 ALYKSTSNKPLMIHHLDECPSQALKVFAENKEIQKLAEOFVLNLVYETTDKHLSP 120
QY 152 DGOYPRIMFVDPSTLYRADITGRYSNRLYAYEPADTALLDMKKALKLKTTEL 206
DB 121 DGOYPRIMFVDPSTLYRADITGRYSNRLYAYEPADTALLDMKKALKLKTTEL 175

RESULT 10

AAB72203
ID AAB72203 standard; Protein: 175 AA.

XX AC AAB72203;

DT 04-MAY-2001 (first entry)

XX DE Human huxAG-1/CCSG colon cancer specific gene amino acid sequence.

XX KM Human; growth factor; huxAG-1; colon cancer specific gene; CCSG;
XX cell proliferation; liver disease; fulminant liver failure; cirrhosis;
XX hepatitis; cancer; colon cancer; colorectal carcinoma.

XX OS Homo sapiens.

XX PN US6171816-B1.

XX PD 09-JAN-2001.

XX PF 22-AUG-1997; 97US-0916576.

XX PR 23-AUG-1996; 96US-0024347.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Yu G, Dillon PJ, Ebner R, Endress GA;

XX DR WPI; 2001-136872/14.

XX DR N-PSDB; AAF63314.

PT Novel human growth factor polypeptide useful for diagnosing and
PT treating colon cancer and liver diseases, to prevent and heal damage to
PT the lungs and for identifying modulators of therapeutic use -

XX PS Claim 165; Fig 1; 59pp; English.

CC This invention relates to a human growth factor polypeptide huxAG-1 also
CC known as a colon cancer specific gene (CCSG). HuxAG-1 stimulates cell
CC proliferation as a growth factor. The huxAG-1 protein is useful for
CC identifying compounds capable of enhancing or inhibiting cellular
CC response induced by huxAG-1. The protein is also useful for stimulating
CC proliferation of cells e.g. colon, breast, liver and lung cells, and
CC hepatocytes. It is useful for alleviating or treating liver diseases and
CC pathologies such as fulminant liver failure caused by cirrhosis, liver
CC damage caused by viral hepatitis and toxic substances, for preventing and
CC treating damage to the lungs caused by various pathological states and
CC for promoting healing of breast tissue injury due to surgery, trauma or

CC cancer. huxag-1 and the identified antagonist are useful for treating
 CC cancer, in particular colon cancer. Detecting altered levels of huxag-1
 CC and its polynucleotides are useful for diagnosing or detecting cancer in
 CC mammals. The gene encoding huxag-1 is useful for monitoring human
 CC colorectal carcinoma. huxag-1 nucleic acid molecules are also useful for
 CC chromosome identification. The present sequence represents the huxag-1
 CC protein.

XX Sequence 175 AA;

Query Match 84.6%; Score 899; DB 22; Length 175;

Best Local Similarity 100.0%; Pred. No. 1.8e-90;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIVSAFLLVALSYTLARDITVKGAKKDKSRPKLPOTLSRGWDDLIWQTYEE 91
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 MEKIVSAFLLVALSYTLARDITVKGAKKDKSRPKLPOTLSRGWDDLIWQTYEE 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 92 ALYKSTSNKPLMIITHHDECPHSQALKKVFENKEIOKLAOFVLNLVYETDKHLS 151
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db .61 ALYKSTSNKPLMIITHHDECPHSQALKKVFENKEIOKLAOFVLNLVYETDKHLS 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 152 DGQYVPRIMFVDPSTLVRADITGRYSNRLYAYEPADFTALLDNMKKALKLTTEL 206
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 DGQYVPRIMFVDPSTLVRADITGRYSNRLYAYEPADFTALLDNMKKALKLTTEL 175
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 11

AAG00109 AAG00109 standard; Protein; 132 AA.

XX AAG00109;

DT 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 4190.

DE Human secreted protein, SEQ ID NO: 4190.

XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping.

XX Homo sapiens.

OS Homo sapiens.

XX EP103401-A2.

XX 06-SEP-2000.

PD 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

PR (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

PI WPI; 2000-500381/45.

XX N-PSDB; AAC00115.

DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 4190; 71pp + CD-ROM; English.

PS The present sequence is a polypeptide encoded by one of a large number

CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs

CC were prepared from total human RNAs or polyA+ RNAs derived from 30

CC different tissues. EST sequences usually correspond mainly to the 3'

CC untranslated region (UTR) of the mRNA because they are often obtained

CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for

CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in

CC those cases where longer cDNA sequences have been obtained, the full 5'

CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'

CC ends and can therefore be used to obtain full length cDNAs and genomic

CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 132 AA;

Query Match 61.7%; Score 656; DB 21; Length 132;

Best Local Similarity 100.0%; Pred. No. 5.9e-64;
 Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIVSAFLLVALSYTLARDITVKGAKKDKSRPKLPOTLSRGWDDLIWQTYEE 91
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 MEKIVSAFLLVALSYTLARDITVKGAKKDKSRPKLPOTLSRGWDDLIWQTYEE 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 92 ALYKSTSNKPLMIITHHDECPHSQALKKVFENKEIOKLAOFVLNLVYETDKHLS 151
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 ALYKSTSNKPLMIITHHDECPHSQALKKVFENKEIOKLAOFVLNLVYETDKHLS 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 152 DGQYV 157
 |||||||
 Db 121 DGQYV 126

RESULT 12

AAV59718 AAV59718 standard; Protein; 131 AA.

XX AAV59718;

DT 18-JAN-2000 (first entry)

XX Secreted protein 78-21-1-B7-FL1.

DE Secreted protein; fingerprint identification technique;

XX chromosome mapping; human; hereditary disease; diagnosis; cancer;

KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;

KW autoimmune disease; rheumatic disease; embryonic disorder; myopathy;

KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;

XX hypertension.

XX Homo sapiens.

OS Homo sapiens.

XX WO940189-A2.

XX 12-AUG-1999.

PD 09-FEB-1999; 99WO-IB00282.

XX 09-FEB-1998; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

PR 10-AUG-1998; 98US-0096116.

PR 04-SEP-1998; 98US-0099273.

XX (GEST) GENSET.

PI Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX WPI; 1999-600966/51.

XX N-PSDB; AA240846.

DR Extended cDNAs useful for expressing secreted proteins and to obtain

XX specific antibodies -

XX Claim 10; Page 240; 244pp; English.

PS This sequence represents a human secreted protein of the invention.

CC The extended cDNAs (or genomic DNAs obtainable from them) may be used to

CC prepare PCR primers and probes. These are useful for forensic matching or

CC positive identification by DNA sequencing. They may also be used in

CC alternative fingerprint identification techniques. Antibodies against the

CC proteins encoded by the extended cDNAs are useful in identification of

CC tissue types or cell species, as well as identifying tissue specific

CC soluble proteins. The sequences can be used for chromosome mapping and

CC Identification of genes associated with hereditary diseases or drug
 CC response: signal sequences from the cDNAs can be used in construction of
 CC secretion vectors. Other sequences derived from the extended cDNAs can be
 CC used to clone upstream genomic DNA sequences including promoters. This is
 CC in turn useful for identifying proteins that interact with promoter
 CC sequences. Some of the proteins may be useful in diagnosing and treating
 CC several disorders including, but not limited to: cancer, hyperlipidaemia,
 CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
 CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
 CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.

XX Sequence 131 AA:

Query Match 56.3%; Score 605; DB 20; Length 131;
 Best Local Similarity 92.9%; Pred. No. 2.3e-58;
 Matches 117; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 32 MEKIPVSAFLILVALSTYLARDTVKPGAKKDTKDSRPKLQPTLSRGMGDLMTQTYEE 91
 Db 1 MEKIPVSAFLILVALSTYLARDTVKPGAKKDTKDSRPKLQPTLSRGMGDLMTQTYEX 60

QY 92 ALYKSTSNKPLMIHHLDECPHSQALKKVAENKEIQKLAEOFLNLVYETTDKHLSP 151
 Db 61 XLKSKTSNKPMLIHHLDKCPHSQALKKVAENKXIQLKLAOFVLNLVYETTDKHLSP 120

QY 152 DGQYVP 157
 Db 121 DGQYXP 126

RESULT 13

AAV12312
 ID AAV12312 standard; Protein: 116 AA.

XX AAV12312;

DT 17-JUN-1999 (first entry)

XX Human 5' EST secreted protein SEQ ID NO:343.

XX Human: secreted protein; EST: expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.

XX WO9906548-A2.

XX 11-FEB-1999.

PF 31-JUL-1998; 98MO-IB01222.

PR 01-AUG-1997; 97US-0905135.

XX (GEST) GENSET.

PI Duciect A, Dumas Milne Edwards J, Lacroix B;

XX WPI; 1999-153778/13.

DR N-PSDB; AAX41145.

PT New nucleic acids encoding human secreted proteins - obtained from
 PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
 PT kidney, lung, umbilical cord, placenta and colon tissue

XX Claim 27; Page 682; 824pp; English.

CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins, and encode the proteins given in AAV12261 to
 CC AAV12314, respectively. The proteins given represent the signal peptide

CC and an N-terminal fragment of a secreted protein. The nucleic acid
 CC sequences can be used for producing secreted human gene products. They
 CC can also be used to develop products for diagnosis and therapy. The
 CC proteins obtained may have cytokine activity, cell
 CC proliferation/differentiation activity, haematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity or other activities. The products
 CC can be used in forensic, gene therapy and chromosome mapping procedures.
 CC The sequences can also be used for obtaining corresponding promoter
 CC sequences. The nucleic acids encoding the signal peptide can be used for
 CC directing extracellular secretion of a polypeptide or the insertion of a
 CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX Sequence 116 AA:

Query Match 56.3%; Score 598; DB 20; Length 116;
 Best Local Similarity 100.0%; Pred. No. 1.2e-57;
 Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPVSAFLILVALSTYLARDTVKPGAKKDTKDSRPKLQPTLSRGMGDLMTQTYEE 91
 Db 1 MEKIPVSAFLILVALSTYLARDTVKPGAKKDTKDSRPKLQPTLSRGMGDLMTQTYEX 60

QY 92 ALYKSTSNKPLMIHHLDECPHSQALKKVAENKEIQKLAEOFLNLVYETTDK 147
 Db 61 ALYKSTSNKPLMIHHLDECPHSQALKKVAENKEIQKLAEOFLNLVYETTDK 116

RESULT 14

AAV64672
 ID AAV64672 standard; Protein: 115 AA.

XX AAV64672;

DT 01-FEB-2000 (first entry)

XX Human 5' EST related polypeptide SEQ ID NO:833.

XX Human: 5' EST: expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location development; protein synthesis; stability;
 KW regulation; identification.

OS Homo sapiens.

XX WO9953051-A2.

XX 21-OCT-1999.

PF 09-APR-1999; 99MO-IB00712.

PR 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duciect A, Giordano J;

XX WPI; 2000-038446/03.

DR N-PSDB; AA242286.

PT Novel secreted protein 5' expressed sequence tag sequences used in
 PT diagnostic, forensic, gene therapy, and chromosome mapping procedures

XX Claim 3; Page 604; 837pp; English.

CC AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAV64651 to
 CC AAV64658 represent the EST-related proteins corresponding to AA242265 to
 CC AA243052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated

CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AA242249 to AA242264 and AA64644 to AA64650 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 115 AA;

Query Match 55.8%; Score 593; DB 21; Length 115;
Best Local Similarity 100.0%; Pred. NO. 4e-57;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 MEKIPVSALLVALSTLTARDITVKGAKKDTKDSRKLPTQLSRGMDLIMTQYEE 91
1 MEKIPVSALLVALSTLTARDITVKGAKKDTKDSRKLPTQLSRGMDLIMTQYEE 60

Db 92 ALYKSTSNKRLMIHHLDECPHSOALKKVAENKEIOKLAEOFYLLMLVYETTD 146
61 ALYKSTSNKRLMIHHLDECPHSOALKKVAENKEIOKLAEOFYLLMLVYETTD 115

RESULT 15
AAW37846
ID AAW37846 standard; Protein: 166 AA.

XX AAW37846;
AC
XX 28-AUG-1998 (first entry)
DE Human XAG growth factor huxAG-3.
XX
XX huxAG-3; XAG; growth factor; breast disease; liver disease;
KM lung disease; emphysema; wound healing; cancer; diagnosis; therapy;
KM human.

XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Peptide 1..23
FT /label- sig-peptide
FT Protein 24..166
FT /label- Mat-protein
FT /note- "Claim 11"
FT Peptide 25..44
FT /label- Epitope
FT /note- "Claim 12"
FT Peptide 104..115
FT /label- Epitope
FT /note- "Claim 12"

XX MO9807749-A1.
PN
XX
XX 26-FEB-1998.
PD
XX 22-AUG-1997; 97WO-US14139.
PF
XX 23-AUG-1996; 96WO-US13766.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Dillon PJ, Ebner R, Endress GA, Yu G;
PI
XX WPI; 1998-169093/15.
DR

DR N-PSDB: AAV19157.
XX
XX New isolated human XAG growth factor(s) - used to develop products
PT for treating e.g. liver, lung or breast diseases or
PT hyperproliferative disorders, e.g. cancer.

XX Claim 1; Fig 3A-B; 141pp; English.

PS
XX This polypeptide comprises huxAG-3, a member of a novel family of
CC human growth factors also including huxAG-1 (see AAW37844) and
CC huxAG-2 (see AAW37845). These proteins share homology with the
CC XAG protein of *Xenopus laevis*, which is involved in embryogenesis
CC and is expressed in adult tissue. huxAG-3 cDNA (see AAV19157) was
CC isolated from a human small intestine cDNA library. Vectors, host
CC cells, antibodies, and screening methods for identifying agonists
CC and antagonists of huxAG-3 are provided. huxAG polypeptides are
CC growth factors and can be used to stimulate proliferation of cells.
CC They can be used to stimulate the proliferation and differentiation
CC of hepatocytes to alleviate or treat liver diseases and pathologies
CC such as fulminant liver failure caused by cirrhosis, liver damage
CC caused by viral hepatitis and toxic substances. They can also be
CC used to stimulate or promote liver regeneration, e.g. after
CC surgery. They can also be used to prevent and heal damage to the
CC lungs caused by various pathological states. They can be used to
CC stimulate proliferation and differentiation and promote the repair
CC of alveoli and bronchiolar epithelium to prevent, attenuate, or
CC treat acute or chronic lung damage, e.g. emphysema, which results
CC in the progressive loss of alveoli, and inhalation injuries, e.g.
CC resulting from smoke inhalation and burns, that cause necrosis of
CC the bronchiolar epithelium and alveoli. They can also be used to
CC stimulate the proliferation and differentiation of breast tissue
CC and could therefore be used to promote healing of breast tissue
CC injury due to surgery, trauma or cancer. Antagonists can be used
CC to treat hyperproliferative disorders, including cancer, in
CC particular hepatocellular carcinoma, osteoclastoma, breast cancer,
CC or colon cancer. The products can also be used for detection and
CC diagnosis.

XX
XX Sequence 166 AA;
SO

Query Match 52.2%; Score 555; DB 19; Length 166;
Best Local Similarity 64.7%; Pred. NO. 1e-52;
Matches 108; Conservative 28; Mismatches 19; Indels 12; Gaps 4;

OY 41 LLLVALSTLTARDITVKGAKKDTKDSRKLPTQLSRGMDLIMTQYEEALYKSTSN 100
11 LLLVTVSSNLA--IAIK-----KEKRP--PQTLSRGMDITWVQYEEGLFYAQSK 59

OY 101 KPLMTIHHLDECPHSOALKKVAENKEIOKLAE-OFVLLNLYVETTDHLSPDGQYVRI 159
60 KPLMTIHHLDECPHSOALKKVAENKEIOKLAE-OFVLLNLYVETTDHLSPDGQYVRI 119

OY 160 MEVDSLTFRADITGRYSNRLYAVEPADTALLDNMKAKLILKTEL 206
120 MEVDSLTFRADITGRYSNRLYAVEPADTALLDNMKAKLILKTEL 166

Search completed: October 9, 2002, 16:40:04
Job time : 33 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:39:22 ; Search time 21 Seconds

(without alignments)
942.591 Million cell updates/sec

Title: US-09-674-266a-181

Perfect score: 1063

Sequence: 1 RLSCAGTLSSGSGPHPSRRLT.....DTALLDNMKKALKLTTEL 206

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	899	84.6	175	2 JE0350	Anterior gradient-
2	199	18.7	186	2 T31643	hypothetical prote
3	119	11.2	150	2 D69100	probable protein d
4	87.5	8.2	458	2 S50919	hypothetical prote
5	87	8.2	348	2 G72016	probable disulfide
6	87	8.2	348	2 C86607	disulfide bond iso
7	85.5	8.0	1830	1 S19188	myosin-V - chicken
8	84.5	7.9	1151	2 AG1717	probable peptidogl
9	84	7.9	771	2 S51421	hypothetical prote
10	83.5	7.9	217	2 T37859	probable transcrip
11	83.5	7.9	918	2 C96829	unknown protein fi
12	83	7.8	479	2 T40683	cell cycle protein
13	83	7.8	507	2 S56143	lipote-protein 1i
14	82.5	7.8	788	2 C70441	inositol polyphosp
15	81.5	7.7	970	2 S29069	probable rhol GDP-
16	81.5	7.7	1158	2 T50454	DNA topoisomerase
17	81.5	7.7	1397	2 T10466	hypothetical prote
18	81	7.6	610	2 T22909	hypothetical prote
19	81	7.6	635	2 T23874	probable protein d
20	80	7.5	259	2 A96740	hypothetical prote
21	80	7.5	259	2 A96740	hypothetical prote
22	80	7.5	843	2 S53947	hypothetical prote
23	79.5	7.5	277	2 T37629	hypothetical prote
24	79.5	7.5	301	2 G71929	ABC transporter (A
25	79.5	7.5	642	2 T83718	hypothetical prote
26	79.5	7.5	921	2 T84593	probable rna polym
27	79	7.4	477	2 T37791	carboxypeptidase C
28	79	7.4	510	1 S43516	shikimate 5-dehydr
29	78.5	7.4	280	2 A69590	

30	78.5	7.4	917	2 T05430	hypothetical prote
31	78.5	7.4	1033	2 A96714	hypothetical prote
32	78.5	7.4	1262	2 I48855	tripeptidyl-peptid
33	78.5	7.4	1570	2 AC2012	hypothetical prote
34	78	7.3	337	2 S06956	segmentation prote
35	78	7.3	1828	2 B59254	myosin heavy chain
36	78	7.3	1853	1 A46761	myosin heavy chain
37	78	7.3	1855	2 A59254	myosin heavy chain
38	76.5	7.2	985	2 E69850	formate dehydrogen
39	76.5	7.2	1025	2 T10259	RNA-directed DNA p
40	76	7.1	166	2 E72018	thioredoxin disulf
41	76	7.1	166	2 D86606	thioredoxin disulf
42	76	7.1	367	2 T39574	probable uracil ki
43	76	7.1	543	1 F64871	oligopeptide-bindi
44	76	7.1	583	2 E75529	probable peptide A
45	76	7.1	1651	2 F88750	protein vlt-6 limp

ALIGNMENTS

RESULT 1
JE0350
Anterior gradient-2 - human
C:Species: Homo sapiens (man)
C>Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C:Accession: JE0350
R:Thompson, D.A.; Weigelt, R.J.
Biochem. Biophys. Res. Commun. 251, 111-116, 1998
A:Title: hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is
A:Reference number: JE0350; MUID:99009231
A:Accession: JE0350
A:Molecule type: mRNA
A:Residues: 1-175 <THO>
A:Cross-References: GB:AF007791; NID:q3779196; PIDN:AAC77358.1; PID:q3779197
C:Comment: This protein is coexpressed with estrogen receptor (ER) in breast cancer c
C:Genetics:
A:Gene: hAG-2
A:Superfamily: human anterior gradient-2

Query Match 84.6%; Score 899; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 3.2e-74;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPVSAPFLLVASYTLARDTYVKGAKKDKOSRPLPOTLSRGWDQILMTQYEE 91
|||||
Db 1 MEKIPVSAPFLLVASYTLARDTYVKGAKKDKOSRPLPOTLSRGWDQILMTQYEE 60

QY 92 ALYKSTSNKPLMIITHHDECPHSQALKVFAENKEIQKLAQFVLNLYVETTKHLS 151
|||||
Db 61 ALYKSTSNKPLMIITHHDECPHSQALKVFAENKEIQKLAQFVLNLYVETTKHLS 120

QY 152 DCOYVPRIMEVPSLTVRADITGRYSNRLYAEPAADTALLDNMKKALKLTTEL 206
|||||
Db 121 DCOYVPRIMEVPSLTVRADITGRYSNRLYAEPAADTALLDNMKKALKLTTEL 175

RESULT 2
T31643
hypothetical protein Y57A10A u - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
C:Accession: T31643
R:Smey, R.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z21048
A:Accession: T31643
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-186 <WIL>
A:Cross-References: EMBL:AL117195; PIDN:CAB55026.1; CESP:Y57A10A.u
A:Experimental source: clone Y57A10A
C:Genetics:


```

Oy 29 RVAMEKIPVSAFLVLLVSTYLRADTVYKPGAKKDTKDSRPRLPQLTSGWGDQLMTQT 88
Db 9 RCSIOLKQKLVATLTL---LSLSL---PTLEAAENRDS-----DSIYWHLD 46

Oy 89 YEEALYKSKTSNKPMLTIHHHDEC--PHSQALKKYFAENKEIOKLAEOFVLNLVY 142
Db 47 YQEALQKSKRAEPLLVIFSGSDMNGPCMKIRKEVLESEFEIKRVGGRVCVEVEY 102

RESULT 6
C66607
disulfide bond isomerase [imported] - Chlamydomophila pneumoniae (strain J138)
C.Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C.Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C.Accession: C66607
R.Shiral, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ishiguro, T.
Nucleic Acids Res. 28, 2311-2314, 2000
A.Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A.Reference number: A86491; MUID:20330349
A.Accession: C66607
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-348 <STO>
A.Cross-references: GB:BA000008; NID:98979307; PIDN:BA09941.1; GSPDB:GN00142
A.Experimental source: strain J138
C.Genetics:
A.Gene: CPj0933

Query Match 8.2%; Score 87; DB 2; Length 348;
Best Local Similarity 25.0%; Pred. No. 3.5;
Matches 29; Conservative 26; Mismatches 37; Indels 24; Gaps 4;

Oy 29 RVAMEKIPVSAFLVLLVSTYLRADTVYKPGAKKDTKDSRPRLPQLTSGWGDQLMTQT 88
Db 9 RCSIOLKQKLVATLTL---LSLSL---PTLEAAENRDS-----DSIYWHLD 46

Oy 89 YEEALYKSKTSNKPMLTIHHHDEC--PHSQALKKYFAENKEIOKLAEOFVLNLVY 142
Db 47 YQEALQKSKRAEPLLVIFSGSDMNGPCMKIRKEVLESEFEIKRVGGRVCVEVEY 102

RESULT 7
S19188
myosin-V - chicken
M.Alternate names: calmodulin-binding protein; myosin I heavy chain, 190K
C.Species: Gallus gallus (chicken)
C.Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Jan-2001
C.Accession: S19188; A44359; B44359; S29249
R.Spreafico, E.M.; Cheney, R.E.; Matteoli, M.; Nascimento, A.A.; De-Camilli, P.V.; Larsen, J.
Submitted to the EMBL data library, February 1992
A.Description: complete cDNA coding sequence of chicken brain p190, a calmodulin binding protein
A.Reference number: S19188
A.Accession: S19188
A.Molecule type: mRNA
A.Residues: 1-1830 <ESP>
A.Cross-references: EMBL:Z11718
R.Spreafico, E.M.; Cheney, R.E.; Matteoli, M.; Nascimento, A.A.; De Camilli, P.V.; Larsen, J.
Cell Biol. 119, 1541-1557, 1992
A.Title: Primary structure and cellular localization of chicken brain myosin-V (p190), a calmodulin-binding protein
A.Reference number: A44359; MUID:93107155
A.Accession: A44359
A.Molecule type: mRNA
A.Residues: 1-1688, 'R', 1690-1830 <ES2>
A.Experimental source: brain
A.Note: sequence extracted from NCBI backbone (NCBIN:121153, NCBIPI:121154)
A.Note: the codon CCG for residue 1689 is inconsistent with the codon AGC for residue 1689
A.Accession: C64459
A.Molecule type: protein
A.Residues: 155-164 <ES3>
R.Sanders, G.; Lichte, B.; Meyer, H.E.; Killmann, M.W.
FEBS Lett. 311, 295-298, 1992
A.Title: cDNA encoding the chicken ortholog of the mouse dilute gene product. Sequence d

```

```

A:Reference number: S29249; MUID:93012002
A:Accession: S29249
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1142,1144-1830 <SAN>
A:Cross-references: EMBL:X67251; NID:g63364; PIDN:CAA47673.1; PID:g63365
C:Comment: The neck domain comprises six approximately 23-residue tandem repeats; thl
C:Superfamily: myosin MYO2; myosin motor domain homolog
C:Keywords: actin binding; ATP; brain; coiled coil; nucleotide binding; P-loop; phospho
F:163-170/Domain: myosin motor domain homology <MOTOR>
F:163-170/Region: nucleotide-binding motif A (P-loop)
F:645-666/Region: actin binding #status predicted
F:765-909/Domain: neck #status predicted <NECK>
F:912-1420/Domain: coiled coil #status predicted <COI>
F:1421-1830/Domain: carboxyl-terminal domain #status predicted <CTD>
F:169/Binding site: ATP (lys) #status predicted
F:1735/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match      8.0%; Score 85.5; DB 1; Length 1830;
Best Local Similarity 28.1%; Pred. No. 38;
Matches 27; Conservative 21; Mismatches 41; Indels 7; Gaps 3;

Qy 63 DTKDSRPRLPOTLSRGWGDOLITQTYEEALY-KKTSNKPLMIHNLDECPH--SQALK 119
Db 507 DLLDECKMPKSDDTMAOKLYNLHLKCALFEKRSLSNKAFFIKHFADKYEYCEGFLE 566
Qy 120 K---VFPAENKEIQKLAEQFVLLNLVEYETDKHLSP 151
Db 567 KKNDIVYEBOIKVLKSSKKFKLLPELPDDEKAIISP 602

RESULT 8
AG1717
probable peptidoglycan bound protein (LPXNG motif) lin2283 [imported] - Listeria inno
C.Species: Listeria innocua
C.Date: 27-Nov-2001 #sequence-revision 27-Nov-2001 #text-change 27-Nov-2001
C.Accession: AG1717
R.Glasner, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgat, O.; Entian, K.D.; Fahlg
.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A.Authors: Kreft, J.; Kunz, M.; Kunst, F.; Krappkat, G.; Madueno, E.; Maicournam, A.;
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlta
A.Title: Comparative genomics of Listeria species.
A.Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AG1717
A.Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1151 <GAS>
A:Cross-references: GB:A1592022; PIDN:CAC97511.1; PID:916414795; GSPDB:GN00178
A.Experimental source: strain Clip11262
C.Genetics:
A:Gene: lin2283

Query Match      7.9%; Score 84.5; DB 2; Length 1151;
Best Local Similarity 22.8%; Pred. No. 26;
Matches 44; Conservative 31; Mismatches 51; Indels 67; Gaps 10;

Qy 50 LARDTTPVRGAK--DTDSRPKLPTQLSRKGODLIQTQ-----TYE 90
Db 961 VALDSVAAPVTQAYDDONHKQIASPETLTGAEGEFTRKOKKITNYTLVKTPANVSSTFN 1020
Qy 91 EA-----LYKSSTNSKPLMIHNLDECPHSQALKVFAENKEIOKLAEQFLV--LNLV 141
Db 1021 EKAQTVTFYQGVNTGCN--IIYDYD-----KNGEKLDSDIVLGKLNSS 1063
Qy 142 YETDKHLSPDQY-VPRIMEVDPSLTVRADITGRYSNR----LYAVEPADTALL----- 191
Db 1064 YRTSAKTISGYLYGTPTK-----NASGRFSMTSGRVFYVEYKASFYSISSNG 1111
Qy 192 LDNNKKALKLKLT 204
Db 1112 ADNVKRSATLPOT 1124

```


A:Gene: SPDB:SPBC776.12c
 A:Map position: 2
 A:Introns: 77/1: 92/2: 331/1
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases: protein kinase homolo

Query Match 7.8%; Score 83; DB 2; Length 479;
 Best Local Similarity 22.4%; Pred. No. 12;

Matches 28; Conservative 24; Mismatches 55; Indels 18; Gaps 5;

OY 83 LIWOTYEALYKSKTSNKP---LMIHHLDECPHSQLKVFANKEIQKLAEQFVLL 138

DB 352 IIMASGASISYKEKLRHPSOEERLCIDPLEKCLELDCKNRISAE---EALDHDFTYL 407

OY 139 -NLVETTDKHLSPDGOYPRIMFVDPSTLVRADITGRSNLYAX---EPADTALLDN 194

DB 408 DMLAEKKDDDTAFDMS-----FGETSPKDEDLTAKHLSHLDKEQDETEPTSLSK 461

OY 195 MKKAL 199

DB 462 RKRST 466

RESULT 13

S56143

cell cycle protein kinase hsk1 (EC 2.7.1.-) - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
 C:Date: 28-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 10-Dec-1999

C:Accession: S56143; T40682

R:Masai, H.; Miyake, T.; Arai, K.

EMBO J. 14, 3094-3104, 1995

A:Title: hsk1(+), a Schizosaccharomyces pombe gene related to Saccharomyces cerevisiae

A:Reference number: S56143; MUID:95347336

A:Accession: S56143

A:Molecule type: DNA

A:Residues: 1-507 <MAS>

A:Cross-references: GB:D50493; NID:9807845; PIDN:BAA09087.1; PID:dl009722; PID:9807846

A:Experimental source: strain j72

R:Lyne, M.; Rajandream, M.A.; Bartell, B.G.; Medler, H.; Kutzner, M.; Wambutt, R.

submitted to the EMBL Data Library, January 1999

A:Reference number: 221944

A:Accession: T40682

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-507 <LVN>

A:Cross-references: EMBL:AL035263; PIDN:CAA22885.1; GSPDB:GN00067; SPDB:SPBC776.12c

A:Experimental source: strain 972h; cosmid c776

C:Genetics:

A:Gene: hsk1; SPBC776.12c

A:Map position: 2

A:Introns: 77/1: 359/1

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases: protein kinase homolo

C:Keywords: phosphotransferase; protein kinase

F:66-404/Domain: protein kinase homology <KIN>

lipoate-protein ligase A - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999

C:Accession: C70441

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666.

A:Accession: C70441

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-788 <AOP>

A:Cross-references: GB:AE000748; NID:q2983960; PIDN:AAC07512.1; PID:q2983965; GB:AE00

A:Experimental source: strain VPS

C:Genetics:

A:Gene: lplA

Query Match 7.8%; Score 82.5; DB 2; Length 788;

Best Local Similarity 23.0%; Pred. No. 25;

Matches 49; Conservative 34; Mismatches 73; Indels 57; Gaps 10;

OY 25 VKSRVAMEKIPYSAFLVLVALSYTLARDTVYKGAKKDKDSR-PKLPQTLR-GW-G 80

DB 585 VQKROAQKRPVRENKKEPIQNYKVPK---VARGQEDM-SRLPIVETSTGIFNG 639

OY 81 DLIWOTYEALYKSKTSNKPRLMIHHLDECPHSQLKVFANKEIQKLAEQFVLL 140

DB 640 OKVSEFALQVAVELLKAKRPLIIVGPL-----VLMAWNEAKKGGIIRLKE 689

OY 141 VVETTDKHLSPDGOYPRIMFVDPSTLVRADITGRSNLYAX---EPADTALLDN 175

DB 690 KIPNLHILPD--YKPKKKVDFSRVDPNPHLSVLGHEHDLTLMIGVCHRTDFTIR 747

OY 176 YSNR-----LYAEPADTALLDNMKK 197

DB 748 MLKKYTDKIVLYCTLYGHPDADVSLAGVNSKK 780

RESULT 15

S29069

inositol polyphosphate-5-phosphatase homolog - human

C:Species: Homo sapiens (man)

C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 01-Dec-2000

C:Accession: S29069; I68621

R:Altjee, O.; Olivios, I.M.; Okabe, I.; Bailey, L.C.; Nelson, D.L.; Lewis, R.A.; McIn

Nature 358, 239-242, 1992

A:Title: The Lowe's oculocerebrorenal syndrome gene encodes a protein highly homologous

A:Reference number: S29069; MUID:92334430

A:Accession: S29069

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-970 <ATT>

A:Leahy, A.M.; Charnas, L.R.; Nusbaum, R.L.

Hum. Mol. Genet. 2, 461-463, 1993

A:Title: Nonsense mutations in the OCRF-1 gene in patients with the oculocerebrorenal

A:Reference number: I34349; MUID:93278398

A:Accession: I68621

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 883-912 <RSS>

A:Cross-references: GB:S62085; NID:9385336; PIDN:AAB26926.1; PID:9385337

C:Genetics:

A:Gene: OCRF-1

Query Match 7.7%; Score 81.5; DB 2; Length 970;

Best Local Similarity 23.7%; Pred. No. 40;

Matches 45; Conservative 32; Mismatches 46; Indels 67; Gaps 13;

OY 55 TYKPGAKKDTKDSRPK--LPQTLSRGMDLIWOTYEAL-YKS-----KTS-NKPLMTI 106

DB 553 TYKYDSKTDKRWDSGKCRVP-----AMCDRIIMRGTNVNLNYSRHEMLKTSQHKVPSAL 607

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:39:17 ; Search time 13 Seconds

(without alignments)
613.556 Million cell updates/sec

Title: US-09-674-266A-181

Perfect score: 1063
Sequence: 1 RLSCAGTLLSGSGPHPSRRRT.....DFALLDNMKKALLKTEL 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	448	42.1	183	XAG_XENLA	P55868 xenopus lae
2	437.5	41.2	185	NP77_XENLA	P55869 xenopus lae
3	87.5	8.2	458	YMS1_YEAST	P42933 saccharomyc
4	85.5	8.0	1829	MYSA_CHICK	002440 gallus gall
5	84	7.9	811	REXL_YEAST	P46743 saccharomyc
6	83.5	7.9	217	YD67_SCHPO	Q10319 schizosacch
7	83	7.8	507	HSK1_SCHPO	P50582 schizosacch
8	81.5	7.7	901	OCRL_HUMAN	Q01968 homo sapien
9	81.5	7.7	1398	TOP2_PLAFK	P41001 plasmodium
10	80	7.5	339	YTH6_RHOER	P43484 rhodococcus
11	80	7.5	843	YMS1_YEAST	005050 saccharomyc
12	79.5	7.5	301	ERA_HELPJ	0921W0 helicobacte
13	79	7.4	510	CBP1_ORYSA	P37890 oryza sativ
14	78.5	7.4	280	AROE_BACSU	P54374 bacillus su
15	78.5	7.4	1262	TPP2_MOUSE	064514 mus musculu
16	78.5	7.4	1828	MYSA_RAT	094yfl3 rattus norv
17	78	7.3	337	HAIR_DROME	P14003 drosophila
18	78	7.3	887	ITH3_RAT	063416 rattus norv
19	78	7.3	1853	MYSA_MOUSE	099104 mus musculu
20	78	7.3	1855	MYSA_HUMAN	Q09411 homo sapien
21	76.5	7.2	1025	PO21_NASVI	003278 nasonia vit
22	76	7.1	543	OPPA_ECOLI	P23843 escherichia
23	76	7.1	2145	US20_CAEEL	090290 caenorhabdi
24	75.5	7.1	580	TREZ_MYCTU	Q10769 mycobacteri
25	75.5	7.1	2318	NTC3_MOUSE	062230 mus musculu
26	75	7.1	1694	SN_MOUSE	094yfl1 mus musculu
27	74.5	7.0	905	ZO3_MOUSE	094yfl1 mus musculu
28	74	7.0	260	TATD_ECOLI	P27859 escherichia
29	74	7.0	986	GM13_RAT	062839 rattus norv
30	74	7.0	1034	POL_HVZCA	P24107 human immun
31	74	7.0	1651	VIT6_CAEEL	P18948 caenorhabdi
32	73.5	6.9	233	YEG7_YEAST	P39983 saccharomyc
33	73.5	6.9	695	TKT_PICST	P34736 pichia stip

ALIGNMENTS

RESULT 1	ID	XAG_XENLA	STANDARD:	PRT:	183 AA.
AC	P55868:				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Putative secreted protein XAG precursor.				
GN	XAG.				
OS	Xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;				
OC	Xenopodidae; Xenopus.				
OX	NCBI_TaxID=8355;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Cement gland;				
RX	MEDLINE=97003224; PubMed=8850563;				
RA	Sive H., Bradley L.;				
RT	"A sticky problem: the Xenopus cement gland as a paradigm for anteroposterior patterning."				
RL	Dev. Dyn. 205:265-280(1996).				
CC	- SUBCELLULAR LOCATION: Secreted (Probable).				
CC	- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN CEMENT GLAND.				
CC	- SIMILARITY: HIGH TO XENOPUS NP77.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: 076752; AAB18819.1; -				
KW	Signal.				
FT	CHAIN 19 183 POTENTIAL.				
FT	SEQUENCE 183 AA: 20475 MW: 95EL2B06FCB80CB8 CR664;				
SO	SEQUENCE				
Query Match	42.1%; Score 448; DB 1; Length 183;				
Best Local Similarity	49.4%; Pred. No. 1e-34;				
Matches	88; Conservative 29; Mismatches 51; Indels 10; Gaps 3;				
OY	39 AFLIVALSITLARDTTPV---GAKKDKDSRKLP-----QTSRGMGDLITQY 89				009214 caenorhabdi
DB	6 SLVCLVLLSALGEAVLKPKKQAGTTDTPKTDDEPAITKGLKIDRKGSEIEVQY 65				09ptus pagrus majo
OY	90 EEALYSKTSNKPIMTIHLLDECPHSQLKVAENKEITOKLA-EQFVLNLVYETDTH 148				P15704 clostridium
DB	66 EESLAKARENKPLWYIHLLDECPYSIALKAFVADRMQKLAQEDFIMNLVHPVADEN 125				003412 caenorhabdi
OY	149 LSPDGOYVRIMEFVDSLTFRADITGRYSNRLYAYEPADTALLDNMKKALKLTEL 206				057741 methanococ
DB	126 GSPDGHVYVRVIFIDPSLTFRSDLKRGYCNKMYADADIPELITNMKAKASPLKTEL 183				P37512 bacillus su

```

RESULT 2
NP77_XENLA          STANDARD:          PRT:          185 AA.
ID   NP77_XENLA
AC   P55869;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Putative secreted protein NP77 precursor.
GN   NP77.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cement gland;
RA   Aberger F., Schueren C., Lepperdinger G., Richter K., Grunz H.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC   -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN CEMENT GLAND
CC   (POTENTIAL).
CC   -1- SIMILARITY: HIGH, TO XENOPUS XAG.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL: U02110; AABA9974.1; -.
KW   Signal.
FT   SIGNAL.          1      18      POTENTIAL.
FT   CHAIN            19     185     PUTATIVE SECRETED PROTEIN NP77.
SQ   SEQUENCE        185 AA; 20442 MW; AE3807C926044509 CRC64;

Query Match          41.2%; Score 437.5; DB 1; Length 185;
Best Local Similarity 53.5%; Pred. No. 9.7e-34;
Matches 84; Conservative 26; Mismatches 46; Indels 1; Gaps 1;

QY   51 ARDTTVGAKKDKTSRPLKPTLSRGWDQLTWQTEBALYKSKTSNKLMIHHL 110
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   29 AATDTGGAASKSEPAVKTKGLKTLDRGWGEDIEMAQTYEGLAKARENKPLV 88
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   111 ECPHSOLAKVFAENKEIOKLA-EQFVLNLVYETTKHLSPOQYPRIMFVPSLTV 169
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   89 DCPYSIALKKAFAVDKMAOKLAQEDFTMLNLVHPVADENOSPDGHVYPKGI 148
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   170 ADITGRSNRLYAVEPADFTALLDNMKKALKLTTEL 206
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   149 SDLKGRYGNKLYADADIDIPELITNMKAKSFLKTEL 185
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
YMS1_YEAST          STANDARD:          PRT:          458 AA.
ID   YMS1_YEAST
AC   P42933;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypochemical 52.2 kDa protein in Sgs1-MRP24 intergenic region.
GN   YMR191W OR YMR646.03.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Pearson D., Bowman S., Bartell B.G., Rajandream M.A.;


```

```

RL   Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RN   [2]
RP   SEQUENCE OF 1-249 FROM N.A.
RC   STRAIN=W303;
RX   MEDLINE=95059068; PubMed=7969174;
RA   Gangloff S., McDonald J.P., Bendixen C., Arthur L., Rothstein R.;
RT   "The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA
    helicase homolog: a potential eukaryotic reverse gyrase.";
RL   Mol. Cell. Biol. 14:8391-8398(1994).
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL: Z47815; CAAB7812.1; -.
DR   EMBL: U22341; AAB60288.1; -.
DR   SGD: S0004803; YMR191W.
KW   Hypochemical protein.
FT   CONFLICT        128     128     FVTW -> Y (IN REF. 2).
FT   CONFLICT        132     135     FVTW -> IALG (IN REF. 2).
SQ   SEQUENCE        458 AA; 52183 MW; C31107DB28DAE33 CRC64;

Query Match          8.2%; Score 87.5; DB 1; Length 458;
Best Local Similarity 20.1%; Pred. No. 1.2;
Matches 55; Conservative 35; Mismatches 74; Indels 109; Gaps 13;

QY   15 PSRRITG-----RWKRSVAMEKIVSAFLILVALS----- 47
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   82 PTRMMAVGGNMWSMLMSVHLRQIRKSLDRILSLSHGNFSHOYRNIFVTWMSLFE 141
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   48 -----YTLARDTVYKGAKKTK-----DSRPKLPQTLSSRG-----W-----GD 81
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   142 ASTAFRRASGILFSPILTRGIGARFDHRRPVNWSKFAFPVRKGAIRGLFTNMNTTSK 201
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   82 QLTWQTEYEALYK-----SKTSNKLPLMT-----IHLDECPHSQ----- 116
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   202 RLQGRAYVTSSTIKFTQEVANNMTISLRCFENSGLGNCSSKCAQONASVNTSKOD 261
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   117 -----ALKV-----FAENKEIOKLAQFVLNLVYETTKHLSPOQYV----- 156
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   262 HVGQVPAKTKSQKDINFRMLEFKIMK--TQNEVDETSAVYMEKPSYIEFTISEFNV 319
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY   157 -----PRIMFVPSLTVRADITGRYSNRLYAYE 184
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   320 NGTFSAPLSFLDPSLL--ADLDEMIRN--YKYE 348
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
MYS4_CHICK          STANDARD:          PRT:          1829 AA.
ID   MYS4_CHICK
AC   002440;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle) (Myosin
GN   MYO5A.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93012002; PubMed=1383040;
RA   Sanders G., Lichte B., Meyer H.E., Kilmann M.W.;
RT   "cDNA encoding the chicken ortholog of the mouse dilute gene product.
    Sequence comparison reveals a myosin I subfamily with conserved C-


```

```

RT      terminal domains." ;
RL      FEBS Lett. 311:295-298(1992).
RN      [2]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 1154-1163.
RC      TISSUE=Brain;
RX      MEDLINE=93107155; Pubmed=1469047;
RA      Espresifco E.M., Cheney R.E., Matteoli M., Nascimento A.A.,
RA      de Camilli P.V., Larson R.E., Mooseker M.S.;
RT      "Primary structure and cellular localization of chicken brain
RT      myosin-V (p190), an unconventional myosin with calmodulin light
RT      chains " ;
RL      J. Cell Biol. 119:1541-1557(1992).
CC      -1- FUNCTION: PROCESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
CC      STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
CC      MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
CC      BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
CC      FORMATION (BY SIMILARITY)
CC      -1- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC      CALMODULIN OR MYOSIN LIGHT CHAINS.
CC      -1- SUBCELLULAR LOCATION: GOLGI-DERIVED CYTOPLASMIC MEMBRANES
CC      (POTENTIAL) .
CC      -1- TISSUE SPECIFICITY: NEURONAL AND NONNEURONAL CELLS OF THE BRAIN.
CC      -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC      -1- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC      -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on ways
CC      use by non-profit institutions as long as their content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).

```

```

CC
DR EMBL; X67251; CAA47673.1; -.
DR EMBL; 211718; CAA77782.1; -.
DR HSSP; P08799; 1MMG.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR ProDom; PD003376; DIL; 1.
DR SMART; SM00015; IQ; 6.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 6.
KW Myosin: Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Phosphorylation.
FT DOMAIN 1 766 MYOSIN HEAD-LIKE.
FT DOMAIN 767 789
FT DOMAIN 790 814 IQ 1.
FT DOMAIN 815 837 IQ 2.
FT DOMAIN 838 862 IQ 3.
FT DOMAIN 863 887 IQ 4.
FT DOMAIN 888 915 IQ 5.
FT DOMAIN 916 1239 IQ 6.
FT DOMAIN 1239 1419 COILED COIL (POTENTIAL).
FT DOMAIN 1419 1766 COILED COIL (POTENTIAL).
FT NP_BIND 1661 1766 DILUTE.
FT NP_BIND 163 170 ATP (BY SIMILARITY).
FT DOMAIN 644 666 ACTIN-BINDING (POTENTIAL).
FT MOD_RES 1734 1734 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 1742 1742 E -> EQ (IN REF. 2).
SQ SEQUENCE 1829 AA; 212381 MW; 0538B278DFC09F6E CRC64;

```

```

Query Match      8.0%: Score 85.5: DB 1: Length 1829;
Best Local Similarity 28.1%: Pred. No. 11;
Matches 27: Conservative 21: Mismatches 41: Indels 7: Gaps 3;

07 63 DTKSDRPLPOLSLGSGMDQLWPTQYEALY-KKTSKNKPLMIITHHDECH--SQALK 119
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```
Db      DLDDECKMPRGSDDTMQIKLYNTHLNKCALFEKRRLSNKAFFIIKHFPADKYEYQCEGPLE 566
Oy      120 K----VFAENKEIQKLAEQFVLNLVYETTDKHLSP 151
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      567 KNKDVTVEEQIKVKLSKKFKFLPELFODEERKAISP 602

RESULT 5
REFL_YEAST
ID       REFL_YEAST    STANDARD;          PRT;           811 AA.
AC       P48743;
DT       01-FEB-1996 (Rel. 33, Created)
DT       15-JUL-1998 (Rel. 36, Last sequence update)
DT       16-OCT-2001 (Rel. 40, Last annotation update)
DE       Hypothetical 90.6 kDa protein in CBP5-DKAI intergenic region.
GN       YLR176C OR L9470.18.
OS       Saccharomyces cerevisiae (Baker's yeast).
OC       Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC       Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX       NCBI_TaxID=4932;
RN       [1]
RP       SEQUENCE FROM N.A.
RC       STRAIN=S288C / AB972;
RA       Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA       Favella A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA       Kubacka T., Hallsworth K., Hawkins J., Hillier L., Je T.,
RA       Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA       Marks E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
RA       Raden L., Riley S., Talch A., Treviskis E., Vignati D.,
RA       Wilcox L., Wolldman P., Vaudin M., Wilson R., Waterson R.;
RL       Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
CC       - SIMILARITY: BELONGS TO THE RXF FAMILY.
-----
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
-----
DR       EMBL; U17246; AAB67470.1; -.
DR       SGD; S0004166; YLR176C.
DR       InterPro; IPR003150; RFX_DNA_binding.
DR       Pfam; PF02257; RFX_DNA_binding; 1.
KW       Hypothetical protein.
SQ       SEQUENCE   811 AA;  90583 MW;  116A88B7DDEAFBF6 CNC64;

Query Match               7.9%; Score 84; DB 1; Length 811;
Best Local Similarity     25.7%; Pred. No. 5.4;
Matches 47; Conservative 29; Mismatches 67; Indels 40; Gaps 8;

Oy      8 LGSGSP-----HPSRLTGCR---WYRSRYAMEKITPYSAFLLVALSYTLAR 52
        ||||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db      129 LGSSPGSHKPSPTPTOPPAOPATROAPATYSVPASISLNSNSSAVPLSKSETTLNN 188

Oy      53 D--TTVK-----PGAKKDPDKSRPKLPQLTSRGMCDQLIWTQTYYEALYKSTKNKP 102
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      189 NPPTAAKTNTNPFTSPSTTKOKOTSOEKRISSISRNRNOETIAKIAIEN--NKSTIEY 246

Oy      103 LMITHH-----LDECPHQALKKVFAENKEIQKLAEQFVLNLVYETTDKHLSPDGQY 155
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      247 AQIVKHAELIKVLSMSQMTSKAAQLDALONRFRER--GVFALLMLMKCKCKSOHDS----Y 300
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy      156 VPR 158
        |||
Db      301 VPR 303

RESULT 6
YD67_SCHPO
ID       YD67_SCHPO    STANDARD;          PRT;           217 AA.
NC       Q10319;
```

```
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U17246; AAB67470.1; -.
DR      SGD; S0004166; YLR176C.
DR      InterPro; IPR003150; RFX_DNA_binding.
DR      Pfam; PF02257; RFX_DNA_binding; 1.
KW      Hypothetical protein.
SQ      SEQUENCE      811 AA; 90583 MW; 116A8BB7DDE4FBF0 CRC64;
Query Match          7.9%; Score 84; DB 1; Length 811;
Best Local Similarity 7.7%; Pred. No. 5.4;
Matches 47; Conservative 29; Mismatches 67; Indels 40; Gaps 8;
QY      8 LSGS-----HPSRRLTOGR----VWRKSRVAMEKRIPVSAPLLVALSYTLAR 52
        | |||| :|: ||: :|:::|:::|:::|:::|:::|:::|
Db      129 LIGSPGSHKSPPTTPQPAOPATRGQPATYSVYPASISLNKSNSSAVPLSKSEETLNN 188
        :|:-|-|-----PGAKDKTKDSRPYLPOTLRSKGWGDQLIWTGYTEBALYKSTSMKP 102
        :|:-|-|-----NPPFAKRTNTFPSPISSTKKOKTSQEKRISSIRRNTOELIAKAIAEN--NKSRTIEEY 246
QY      QY      103 LMIIHH-----LDECHSOALKKVFAFNKEITQLAEQFYLLNLVYETTDKHLSPPDOY 155
        :|:-|-|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      247 AQIVKHAIEIKVLSDMSQNTSKAAIQALAOAONREHER--GVFALLMLMKCKNSOHDSE---Y 300
QY      QY      156 VPR 158
        |||
Db      Db      301 VPR 303

RESULT 6
YD67_SCHPO
ID      YD67_SCHPO      STANDARD:      PRT;      217 AA.
AC      Q10319;
```

```

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 24.9 kDa protein C17G8.07 in chromosome I.
GN SPAC17G8.07.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO HUMAN AFG AND ENL AND YEAST TRANSCRIPTION
CC INITIATION FACTOR TFIIF SMALL SUBUNIT (TFI03/ANCL).
CC -1- SIMILARITY: STRONG, TO YEAST YNL107W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z69795; CAA93690.1; -.
DR Hypothetical protein.
KM SEQUENCE 217 AA; 24946 MW; A274C2AFB0620D7F CRC64;
SQ
Query Match 7.9%; Score 83.5; DB 1; Length 217;
Best Local Similarity 20.6%; Pred. No. 1.1;
Matches 42; Conservative 33; Mismatches 80; Indels 49; Gaps 8;
QY 14 HPSRLTQG-----RWKRSRAAMEKIPYSAFLVLAALSTYARDTYVPGKKKTD 66
DB 39 HTWIFVEGVGDEDISKVRK-----VVEFKL-----HDTYNNP---TRTIE 76
QY 67 SRPLPQTLSSGMDQLIMPTQYEALYKSKTNSKPLMIHHIDECPSQALKVFAENK 126
DB 77 SPPF--EVIENGKEPDIWIRIF-----FAPAHKEALFTYHNLHLHYGPMMEKKASGC 130
QY 127 EIQLAEQVLLNLVETTDKHLSP---DGQVPRIMEVDPSSLVRADITGRYSNRLYA 182
DB 131 LVESVOYEIEIYFNPEFYETKLSQNIIGDGH-----GLAVSEPDHPPSQOLEQ 180
QY 183 YEPADTALLDNMKKALKLKTLEL 206
DB 181 DEADKLDPALIOEVAKTTEMKQOV 204

```

RESULT 7

```

HSK1_SCHPO
ID HSK1_SCHPO STANDARD; PRT; 507 AA.
AC P50582; O94678;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell cycle protein kinase hsk1 (EC 2.7.1.-).
GN HSK1 OR SPBC776.12C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC STRAIN=JY2;
RX MEDLINE=95347336; Pubmed=7621824;
RA Masai H., Miyake T., Arai K.-I.;
RT "hsk1", a Schizosaccharomyces pombe gene related to Saccharomyces
RT cerevisiae CDC7, is required for chromosomal replication."

```

```

RL EMBL J. 14:3094-3104(1995).
RN [2]
RP SEQUENCE FROM N.A. (LONG AND SHORT FORM).
RC STRAIN=972;
RA Lyne M., Rajandream M.A., Barrell B.G., Medler H., Kutznar M.,
RA Wambutt R.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR CHROMOSOMAL REPLICATION.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CDC7 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D50493; BAA09087.1; -.
DR EMBL: AL035263; CAA22885.1; -.
DR EMBL: AL035263; CAB38417.1; -.
DR HSSP: Q16539; IWFC.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam: PF00069; Pkinase_3.
DR DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KM Cell cycle: Cell division; Serine/threonine-protein kinase;
KM Transferrase; ATP-binding; Alternative splicing.
FT DOMAIN 68 433 PROTEIN KINASE.
FT NP_BIND 74 82 ATP (BY SIMILARITY).
FT BINDING 129 129 ATP (BY SIMILARITY).
FT ACT_SITE 216 216 BY SIMILARITY.
FT VARSPLOC 92 119 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 507 AA; 58407 MW; 04970E58218441EE CRC64;

```

Query Match 7.8%; Score 83; DB 1; Length 507;

Best Local Similarity 22.4%; Pred. No. 3.7; Mismatches 55; Indels 18; Gaps 5;

Matches 28; Conservative 24; Mismatches 55; Indels 18; Gaps 5;

```

QY 83 LIMPQTEYEALYKSKTSNKP-----LMTIHHIDECPSQALKVFAENKEIQKLAQFVLL 138
DB 380 ILMASCSASITKKEKLLHKKRQSEERLCIDPLEKLELDCKNRISAE---EALDHPFLYL 435
QY 139 -NLVETTDKHLSPDGQVPRIMEVDPSSLVRADITGRYSNRLYA---EPADTALLDN 194
DB 436 DNLAYEKKDDDTAFDNS-----FGETSFEKDEDLTAKHLSHIIDFKQDEETDEPTLSK 489
QY 195 MKKAL 199
DB 490 KRKSI 494

```

RESULT 8

```

OCRL_HUMAN
ID OCRL_HUMAN STANDARD; PRT; 901 AA.
AC Q01966; Q15684; O60800; Q90J55; Q15774; Q90MA5;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Inositol polyphosphate 5-phosphatase OCRL-1 (EC 3.1.3.-) (Lowe's
DE oculocerebrorenal syndrome protein).
GN OCRL OR OCRL1 OR INP55.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).

```


RC TISSUE=Kidney;
 RX MEDLINE=92334430; PubMed=1321346;
 RA Attree O., Olivos I.M., Okabe I., Bailey L.C., Nelson D.L.,
 RA Lewis R.A., McInnes R.R., Nussbaum R.L.;
 RT "The Lowe's oculocerebrorenal syndrome gene encodes a protein highly
 RT homologous to inositol polyphosphate-5-phosphatase.";
 RL Nature 358:239-242(1992).
 RN [12]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Brain;
 RX MEDLINE=97201100; PubMed=9048911;
 RA Nussbaum R.L., Orrison B.M., Janne P.A., Charnas L., Chihault A.C.;
 RT "Physical mapping and genomic structure of the Lowe syndrome gene
 RT OCR1.";
 RL Hum. Genet. 99:145-150(1997).
 RN [13]
 RP SEQUENCE OF 81-901 FROM N.A. (ISOFORMS A AND B).
 RA Pavitt R.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [14]
 RP SEQUENCE OF 426-901 FROM N.A. (ISOFORM B).
 RA Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 RN [15]
 RP SEQUENCE OF 814-843 FROM N.A.
 RX MEDLINE=93278398; PubMed=8504307;
 RA Leahy A.M., Charnas L.R., Nussbaum R.L.;
 RT "Nonsense mutations in the OCR1-1 gene in patients with the
 RT oculocerebrorenal syndrome of Lowe.";
 RL Hum. Mol. Genet. 2:461-463(1993).
 RN [6]
 RP CHARACTERIZATION.
 RX MEDLINE=95281554; PubMed=7761412;
 RA Zhang X., Jefferson A.B., Auehayeckat V., Majerus P.W.;
 RT "The protein deficient in Lowe syndrome is a
 RT phosphatidylinositol-4,5-bisphosphate 5-phosphatase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:4853-4856(1995).
 RN [7]
 RP CHARACTERIZATION.
 RX MEDLINE=98104142; PubMed=9430598;
 RA Zhang X., Hartz P.A., Philip E., Racusen L.C., Majerus P.W.;
 RT "Cell lines from kidney proximal tubules of a patient with Lowe
 RT syndrome lack OCR1 inositol polyphosphate 5-phosphatase and
 RT accumulate phosphatidylinositol 4,5-bisphosphate.";
 RL J. Biol. Chem. 273:1574-1582(1998).
 RN [8]
 RP VARIANTS LS T-367 DEL; G-451; S-463 AND R-524.
 RX MEDLINE=97342912; PubMed=9195559;
 RA Lin T., Orrison B.M., Leahy A.-M., Suchy S.F., Bernard D.J.,
 RA Lewis R.A., Nussbaum R.L.;
 RT "Spectrum of mutations in the OCR1 gene in the Lowe oculocerebrorenal
 RT syndrome.";
 RL Am. J. Hum. Genet. 60:1384-1388(1997).
 RN [9]
 RP VARIANTS LS TYR-375; GLN-500; ASP-508 AND CYS-513.
 RX MEDLINE=98347141; PubMed=9682219;
 RA Lin T., Orrison B.M., Suchy S.F., Lewis R.A., Nussbaum R.L.;
 RT "Mutations are not uniformly distributed throughout the OCR1 gene in
 RT Lowe syndrome patients.";
 RL Mol. Genet. Metab. 64:58-61(1998).
 RN [10]
 RP VARIANTS LS GLN-500 AND GLN-524.
 RX MEDLINE=98293952; PubMed=9632163;
 RA Kawano T., Indo Y., Nakazato H., Shinadzu M., Matsuda I.;
 RT "Oculocerebrorenal syndrome of Lowe: three mutations in the OCR1 gene
 RT derived from three patients with different phenotypes.";
 RL Am. J. Med. Genet. 77:348-355(1998).
 RN [11]
 RP VARIANT LS ARG-522.
 RX MEDLINE=99002770; PubMed=9788721;
 RA Kubota T., Sakurai A., Arikawa K., Shimazu M., Wakui K., Furuhata K.,
 RA Fukushima Y.;
 RT "Identification of two novel mutations in the OCR1 gene in Japanese

RT families with Lowe syndrome.";
 RL Clin. Genet. 54:199-202(1998).
 RN [12]
 RP VARIANTS LS GLU-357; GLU-421; ASP-424 AND TYR-498.
 RX MEDLINE=20382695; PubMed=10923037;
 RA Monnier N., Saire V., Lerouge E., Berthoin F., Lunardi J.;
 RT "OCR1 mutation analysis in French Lowe syndrome patients:
 RT implications for molecular diagnosis strategy and genetic
 RT counseling.";
 RL Hum. Mutat. 16:157-165(2000).
 RN [13]
 RP VARIANTS LS PRO-337 AND GLY-372.
 RX MEDLINE=20232168; PubMed=10767176;
 RA Roschinger W., Muntau A.C., Rudolph G., Roscher A.A.,
 RA Kammerer S.;
 RT "Carrier assessment in families with Lowe oculocerebrorenal syndrome:
 RT novel mutations in the OCR1 gene and correlation of direct DNA
 RT diagnosis with ocular examination.";
 RL Mol. Genet. Metab. 69:213-222(2000).
 CC -1- FUNCTION: CONVERTS PHOSPHATIDYLINOSITOL 4,5-BISPHOSPHATE TO
 CC PHOSPHATIDYLINOSITOL 4-PHOSPHATE. ALSO CONVERTS INOSITOL 1,4,5-
 CC TRISPHOSPHATE TO INOSITOL 1,4-BISPHOSPHATE AND INOSITOL 1,3,4,5-
 CC TETRAPHOSPHATE TO INOSITOL 1,3,4-TRISPHOSPHATE. MAY FUNCTION IN
 CC LYSOSOMAL MEMBRANE TRAFFICKING BY REGULATING THE SPECIFIC POOL OF
 CC PHOSPHATIDYLINOSITOL 4,5-BISPHOSPHATE THAT IS ASSOCIATED WITH
 CC LYSOSOMES.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A (SHOWN HERE) AND B; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: BRAIN, SKELETAL MUSCLE, HEART, KIDNEY,
 CC LUNG, PLACENTA, AND FIBROBLASTS.
 CC -1- DISEASE: DEFECTS IN OCR1 ARE THE CAUSE OF LOWE'S OCULOCEREBRORENAL
 CC SYNDROME, A DISEASE CHARACTERIZED BY HYDROPHALMIA, CATARACT,
 CC MENTAL RETARDATION, VITAMIN D-RESISTANT RICKETS, AMINOACIDURIA,
 CC AND REDUCED AMMONIA PRODUCTION BY THE KIDNEY.
 CC -1- SIMILARITY: BELONGS TO THE INOSITOL-1,4,5-TRISPHOSPHATE
 CC 5-PHOSPHATASE TYPE II FAMILY.
 CC -1- SIMILARITY: CONTAINS 11 GAP DOMAIN.
 CC -1- DATABASE: NAME=Lowe syndrome mutation database;
 CC URL=http://www.nhgri.nih.gov/DIR/GBRB/Lowe/ocr11_mut_db.html.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M88162; AAA59964.1; ALT_INTL.
 DR EMBL; U57627; AAB03839.2; -.
 DR EMBL; AL022162; CAA18150.1; -.
 DR EMBL; AL022162; CAA18151.1; -.
 DR EMBL; Z73496; CAA97842.1; -.
 DR EMBL; S62085; AAB26926.1; -.
 DR MIM; 309000; -.
 DR InterPro: IPR000300; IPCC.
 DR InterPro: IPR00198; RhogAP.
 DR Pfam: PF00783; IPCC; 1.
 DR Pfam: PF00620; RhogAP; 1.
 DR SMART; SM00128; IPCC; 1.
 DR SMART; SM00324; RhogAP; 1.
 KW Hydrolase; Alternative splicing; Disease mutation.
 FT DOMAIN 735 875
 FT VARSPIC 707 714
 FT VARIANT 337 337
 FT VARIANT 357 357
 FT VARIANT 367 367
 FT VARIANT 372 372
 FT VARIANT 375 375
 FT MISSING (IN ISOFORM B).
 FT R->P (IN LOWE SYNDROME).
 FT /FTID=VAR_010169.
 FT G->E (IN LOWE SYNDROME).
 FT /FTID=VAR_010170.
 FT MISSING (IN LOWE SYNDROME).
 FT /FTID=VAR_010171.
 FT V->G (IN LOWE SYNDROME).
 FT /FTID=VAR_010172.
 FT H->Y (IN LOWE SYNDROME).
 FT VARIANT 375 375

CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Corynebacteriaceae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN
RP
SEQUENCE FROM N.A.
CC STRAIN-N186/21;
RX MEDLINE=96082876; PubMed=7583123;
RA Tamura T., Nagy I., Lupas A., Lottspeltch F., Cejka Z., Schoofs G.,
Tanaka K., de Mot R., Baumeister W.,
RT "The first characterization of a eubacterial proteasome: the 20S
RT complex of Rhodococcus."
RL Curr. Biol. 5:766-774(1995).
RN
RP
SEQUENCE OF 1-333 FROM N.A.
CC STRAIN-N186/21;
RX MEDLINE=95138028; PubMed=7836301;
RA Nagy I., Schoofs G., Compernelle F., Proost P., Vanderleyden J.,
de Mot R.,
RT "Degradation of the thiocarbamate herbicide EPC (S-ethyl
RT dipropylcarbamothioate) and biosurfactant by Rhodococcus sp. strain
RT N186/21 involve an inducible cytochrome P-450 system and aldehyde
RT dehydrogenase."
RL J. Bacteriol. 177:676-687(1995).
CC
CC -I- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV2097C AND M.LEPRAE
CC MLCB2533.24.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC
CC EMBL: U26421; AAC5738.1; -
DR EMBL: U17130; AAC45747.1; -
DR InterPro: IPR004347; DUF245.
DR Pfam: PF03136; DUF245; 1.
KW Hypothetical protein.
SQ
SEQUENCE 339 AA; 37462 MW; 23CB953FB8739599 CRC64;

Query Match 7.5%; Score 80; DB 1; Length 339;
Best Local Similarity 29.5%; Pred. No. 4.2;
Matches 31; Conservative 20; Mismatches 36; Indels 18; Gaps 5;

QY 101 KRLMIHHLDECPHSQLKRYAENKEIKLAQFVLNLVETTDKHLSPGQIVPRIM 160
DB 122 REVTAVHNHSHDPTLRA-TVALADGRELGTALQRIYIDRV---DKFMSAGNDPRV- 175
QY 161 FVDPSLTVRADITGRVSNRLVAYE--PADTALLDNMKALKILK 203
DB 176 -----ADLEKMWVLDLIERDPMCAHLDPKAK-LKIL 210

RESULT 11
YMS1_YEAST STANDARD; PRT; 843 AA.
AC 005050.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 93.3 kDa protein in TAP42-CYK2 intergenic region.
GN YMR031C OR YMR973.0AC.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN
RP
SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972.
RA Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO YEAST YKL050C.

CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC
CC EMBL: Z49213; CAA89146.1; -
DR SGI: S0004633; YMR031C.
KW Hypothetical protein.
SQ
SEQUENCE 843 AA; 93345 MW; 0068D65A229B941D CRC64;

Query Match 7.5%; Score 80; DB 1; Length 843;
Best Local Similarity 21.3%; Pred. No. 13;
Matches 38; Conservative 29; Mismatches 57; Indels 54; Gaps 7;

QY 20 TQGRWRKSRVAMERKIPVAFILVA-----LSYTLARDTYTPGAKKDKTS 67
DB 453 TANRHQTEKRTSOEKIKAS-FDALVARMPTKVAERETLEDTKSEIERKKMQQDEKDE 511
QY 68 RPKLPQTLSSRGMDOLIMTQTEALYKSTSNKPLMIHHLDECPHSQLKRYAENKEIKLA 124
DB 512 KARLDQDLEE-----WKKCKEQDITEARKQDELLKRPYH-DLALNAEAHEKTLVEERD 563
QY 125 --NKEIQKIAEQFV-----LINL--VYETTDKHLSPD 152
DB 564 EINAETSRLQDAIVHKKRISGYGNDLDAQNNRIREDKLTLEGTQKTESLSHIND 621

RESULT 12
ERA_HELPJ STANDARD; PRT; 301 AA.
AC 09ZLMO;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE GTP-binding protein era homolog.
GN ERA OR JHP0466.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX Helicobacter.
OX NCBI_TaxID=85963;
RN
RP
SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonghe B.L., Carmel G.,
RA Tumblin P.J., Caruso A., Ulla-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC
CC -I- FUNCTION: BINDS BOTH GDP AND GTP. HAS AN INTRINSIC GTPASE ACTIVITY
CC AND IS ESSENTIAL FOR CELL GROWTH (BY SIMILARITY).
CC PROTEINS. ERA SUBFAMILY.
CC -I- SIMILARITY: CONTAINS 1 KH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC
CC EMBL: AE001480; AAD06035.1; -
DR HSSP: P06616; IEQA.
DR InterPro: IPR004087; KH.
DR InterPro: IPR004044; KH_TYPE_2.

DR Pfam: PF00013; KH-domain; 1.
 DR PROSITE; PS50823; KH_type_2; 1.
 KW GTP-binding; RNA-binding; Complete proteome.
 FT NP_BIND 12 19 GTP (POTENTIAL).
 FT NP_BIND 64 68 GTP (POTENTIAL).
 FT NP_BIND 122 125 GTP (POTENTIAL).
 FT DOMAIN 204 280 KH.
 SQ SEQUENCE 301 AA; 34485 MW; F668C7730882EDBF CRC64;
 Query Match 7.5%; Score 79.5; DB 1; Length 301;
 Best Local Similarity 27.8%; Pred. No. 4;
 Matches 25; Conservative 13; Mismatches 33; Indels 19; Gaps 4;
 QY 89 YEEALYKSKTSNKP-LMIHHDECPHSQALKVFENKEIOKLAEQVL----- 137
 DB 106 YEEFL---SLCQKPHILASKIDATRHQVQLKLT---QEQKYSQSFALVPLSAKKSQ 158
 QY 138 -LNLVYETTDKHLSPDQGVYPRIMEVDPDL 166
 DB 159 NLNALLECISKHLSPSAMLEKDLMSDEKM 188
 RESULT 13
 CBPL_ORYSA STANDARD; PRT; 510 AA.
 ID CBPL_ORYSA
 AC P37890;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Serine carboxypeptidase I precursor (EC 3.4.16.5) (Carboxypeptidase C).
 GN CBPL.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV_YUKIHIKARI;
 RX MEDLINE=94213891; PubMed=8161571;
 RA Washio K., Ishikawa K.;
 RT Cloning and sequencing of the gene for type I carboxypeptidase in rice.
 RL Biochim. Biophys. Acta 1199:311-314(1994).
 CC CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a broad specificity.
 CC -1- PPM: THREE DISULFIDE BONDS ARE PRESENT (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10; ALSO KNOWN AS THE SERINE CARBOXYPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D17586; BAA04510.1; -.
 DR EMBL; D17586; BAA04510.1; -.
 DR PIR; S43516; S43516.
 DR HSSP; P08819; 1WHT.
 DR MEROPS; S10.001; -.
 DR InterPro: IPR000379; Est_11p_thioest_actsite.
 DR InterPro: IPR001563; Serine carboxypeptidase.
 DR Pfam: PF00450; serine_carboxypeptidase; 1.
 DR PRINTS; PR00724; CARBOXYPEPT_SER_SER; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 DR PROSITE; PS00342; MICROBODIES_CTER; 1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 KW Hydroxylase; Carboxypeptidase; Glycoprotein; zymogen; signal.
 FT SIGNAL 1 25
 FT PROPEP 26 36 POTENTIAL.

FT CHAIN 37 510 SERINE CARBOXYPEPTIDASE I.
 FT ACT_SITE 194 194 BY SIMILARITY.
 FT ACT_SITE 434 434 BY SIMILARITY.
 FT ACT_SITE 487 487 BY SIMILARITY.
 FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SITE 508 510 MICROBODY TARGETING SIGNAL (POTENTIAL).
 SQ SEQUENCE 510 AA; 55709 MW; 28896247FA1371CF CRC64;
 Query Match 7.4%; Score 79; DB 1; Length 510;
 Best Local Similarity 37.5%; Pred. No. 8.7;
 Matches 30; Conservative 5; Mismatches 33; Indels 12; Gaps 4;
 QY 73 QTLRSG--WGDOLIMTQYEEALYKSKTSNKP-LMIHHDECPHSQALKVFENKEIOK 130
 DB 259 QTACHGNYWMTT---TDKCNALYKVDTSINDINLYDIEPCYHSRTIKKVTGAN--TK 312
 QY 131 LAEQFVILNLVYETTDKHL 150
 DB 313 LPRKSFQHLG---TTTKPLA 328
 RESULT 14
 AROE_BACSU STANDARD; PRT; 280 AA.
 ID AROE_BACSU
 AC P54374;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Shikimate 5-dehydrogenase (EC 1.1.1.25).
 GN AROD.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / JH642;
 RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S., Sato T., Takeuchi M.;
 RT Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC CATALYTIC ACTIVITY: Shikimate + NADP(+) = 5-dehydroshikimate + NADPH.
 CC -1- PATHWAY: FOURTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
 CC -1- SIMILARITY: BELONGS TO THE SHIKIMATE DEHYDROGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D84432; BAA12445.1; -.
 DR EMBL; D84432; BAA12445.1; -.
 DR Subtilist; BG11522; aROD.
 DR InterPro: IPR002907; Shikimate_DH.
 DR Pfam: PF01488; Shikimate_DH; 1.
 KW Aromatic amino acid biosynthesis; Oxidoreductase; NADP; Complete proteome.
 SQ SEQUENCE 280 AA; 30642 MW; 1EE0CE9C2EDA533F CRC64;
 Query Match 7.4%; Score 78.5; DB 1; Length 280;
 Best Local Similarity 24.8%; Pred. No. 4.5;
 Matches 39; Conservative 17; Mismatches 58; Indels 43; Gaps 5;
 QY 80 GDOLIMTQYEEALYKSKTSNKP-LMIHHDECPHSQALKVFENKEIOK 130
 DB 93 GDKLVGNTDQEGFVKSLMKVLDKRISELFLMIGAGAAARFPTFVRNTPKKFDICNR 152

OY 115 SOALKKVAE-----NKE-----IOKLAEOFVLLNLVYETTDKHLSPDGOVPRIMEVDP 165
DB 153 TLEKAKRLTEATPSFHNKEVUSIKAEERLEQYDVIHTTSVGMTPNVDVPLSLQRAAS 212
OY 166 LTVRADITGRYSNRLAYEPADFTALLDNMKKALKL 202
DB 213 SAVVCDI-----YVNPLOTALKEASQKGLKTL 240

RESULT 15
TPP2_MOUSE
ID TPP2_MOUSE STANDARD: PRT: 1262 AA.
AC 064514:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tripeptidyl-peptidase II (EC 3.4.14.10) (TPP-II) (Tripeptidyl
aminopeptidase).
GN TPP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEADEN X AL;
RX MEDLINE=95091686; PubMed=7998988;
RA Tomkinson B.;
RT "Characterization of cDNA for murine tripeptidyl-peptidase II reveals
alternative splicing."
RL Blochem. J. 304:517-523(1994).
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal tripeptide from a
polypeptide.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
SUBTILASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X81323; CAA57103.1; -
DR HSSP; P00782; 2SPT.
DR MEROPS; S08.090; -
DR MGD; MGI:102724; TPP2.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolyase; Alternative splicing.
FT ACT SITE 44 44 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 264 264 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 449 449 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT VARSPLIC 985 997 MISSING (IN SHORT ISOFORM).
SO SEQUENCE 1262 AA; 139878 MW; D50D22C85344B034 CRC64;

Query Match 7.4%; Score 78.5; DB 1; Length 1262;
Best Local Similarity 22.9%; Pred. No. 31;
Matches 44; Conservative 31; Mismatches 78; Indels 39; Gaps 7;

OY 5 ACTLSGSGPHPSRRLTQGHWRKRSRYAMEKIPVSAPLLIVALSITLARDTYKPGAK--- 61
DB 970 AGSLTLSTKETELKAGOSAKRQGFKKVDIVHYLL-----IPPTKIKNGSKDKE 1021
OY 62 -----KDTKDSRPLPQTLISRGWGDQLIWQTYEALYKSKTSNKLMI--IHHLDECP 113

DB 1022 KDSKEDLKEFTFTALDCLKIQMTTKLSDSDIYNE-LKETYPALPLPYARLHLD--- 1077
OY 114 HSQALKKVAENKEIOKLAEOFVLLNLVYETTDKHLSPDGOVPRIMEVDP---SLTVRA 170
DB 1078 -----AEKERMKRLNEIVDANNAVLI-----SHIDOTALAVYIAMKTDPRPDATIKN 1124
OY 171 DITGRYSNRLA 182
DB 1125 DMDKQKSTLIDA 1136

Search completed: October 9, 2002, 16:40:31
Job time : 15 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:39:22 ; Search time 27 Seconds
(without alignments)
1319.887 Million cell updates/sec

Title: US-09-674-266A-181

Perfect score: 1063
Sequence: 1 RLSCAGTLLSGGPHSRRLT.....DTALLDMKKALKLKTREL 206

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mbc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	899	84.6	175	4	095994
2	828	77.9	175	11	088312
3	548	51.6	159	13	090Y05
4	447	42.1	185	13	042251
5	227	21.4	172	4	096H50
6	224.5	20.1	170	11	09C000
7	222	20.9	172	4	095881
8	199	18.7	186	5	09NA78
9	165	15.5	257	5	09N5S7
10	119	11.2	150	17	027777
11	89	8.4	950	11	0924S5
12	88.5	8.3	737	2	09AF04
13	87	8.2	348	16	09Z6X3
14	84.5	7.9	1151	16	0929J2
15	84	7.9	435	5	09BJK8
16	83.5	7.9	918	10	09CAB8

17	83.5	7.9	1011	5	09U9P6	09u9p6 drosophila
18	83.5	7.9	1409	5	09Y3A7	09y3a7 drosophila
19	83.5	7.9	1409	5	09U6I2	09u6i2 drosophila
20	83	7.8	786	4	09H747	09h747 homo sapien
21	83	7.8	976	5	09U158	09u158 leishmania
22	82.5	7.8	788	16	067557	067557 aquifex aeo
23	81.5	7.7	649	10	09C5B1	09c5b1 arabidopsis
24	81.5	7.7	1158	3	09URR5	09urr5 schizosacch
25	81	7.6	610	5	09XVM2	09xvm2 caenorhabdi
26	81	7.6	635	5	009598	009598 caenorhabdi
27	80.5	7.6	875	12	090G69	090g69 apple.mosai
28	80	7.5	154	16	09PCE7	09pce7 xyella fas
29	80	7.5	259	10	09M9H0	09m9h0 arabidopsis
30	80	7.5	623	10	09FPR8	09fpr8 arabidopsis
31	80	7.5	861	3	09MKX3	09mkx3 bombyx mori
32	79.5	7.5	277	3	013703	013703 schizosacch
33	79.5	7.5	536	10	093V17	093v17 arabidopsis
34	79.5	7.5	642	16	09KFD1	09kfd1 bacillus ha
35	79.5	7.5	921	10	09SKT6	09skt6 arabidopsis
36	79	7.4	275	2	087445	087445 brachyspira
37	79	7.4	337	5	095NH3	095nh3 drosophila
38	79	7.4	477	3	013745	013745 schizosacch
39	78.5	7.4	480	5	09YMD8	09ymd8 drosophila
40	78.5	7.4	302	11	0922K4	0922k4 mus musculu
41	78.5	7.4	339	6	095LR8	095lr8 macaca fasc
42	78.5	7.4	917	10	049498	049498 arabidopsis
43	78.5	7.4	1033	10	09CAK7	09cak7 arabidopsis
44	78	7.3	127	11	09CTG5	09ctg5 mus musculu
45	78	7.3	337	5	095UA9	095ua9 drosophila

ALIGNMENTS

RESULT 1

095994 PRELIMINARY: PRT: 175 AA.

AC 095994: 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SECRETED CEMENT GLAND PROTEIN XAG-2 HOMOLOG (ANTERIOR GRADIENT 2 (XENOPUS LAEVIS) HOMOLOG).
GN XAG-2/R OR XAG-2/C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BREAST;
RX MEDLINE=99009231; PubMed=9790916;
RA Thompson D.A., Weigel R.J.;
RT "XAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines."
RT Biochem. Biophys. Res. Commun. 251:111-116(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [7]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [8]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [9]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [10]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [11]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [12]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [13]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [14]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [15]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [16]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [17]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [18]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [19]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [20]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [21]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [22]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [23]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [24]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [25]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [26]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [27]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [28]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [29]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [30]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [31]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [32]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [33]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [34]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [35]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [36]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [37]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [38]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [39]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [40]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [41]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [42]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [43]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [44]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [45]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [46]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [47]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [48]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [49]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.
RN [50]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RX Strausberg R.;
RT Submitted (Oct-2001) to the EMBL/GenBank/DBD databases.

Query Match 84.6%; Score 899; DB 4; Length 175;
 Best Local Similarity 100.0%; Pred. No. 4.1e-80;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPVSAPFLILVALSYTLARDTTPVPGAKKDKDSRPKLPQTLSRGWGQDLMTQTYEE 91
 1 MEKIPVSAPFLILVALSYTLARDTTPVPGAKKDKDSRPKLPQTLSRGWGQDLMTQTYEE 60

DB 92 ALYKSTSNKPLMIHHLDECPSHQAALKVFAENKEIQKLAEOQVFLNLVYETTDKHLSP 151
 61 ALYKSTSNKPLMIHHLDECPSHQAALKVFAENKEIQKLAEOQVFLNLVYETTDKHLSP 120

QY 152 DGOYVPRIMEFVPSLTVRADITGRYSNRLYAYEPADTALLDMKKALKLKTTEL 206
 121 DGOYVPRIMEFVPSLTVRADITGRYSNRLYAYEPADTALLDMKKALKLKTTEL 175

RESULT 2
 ID 088312 PRELIMINARY; PRT; 175 AA.
 AC 088312;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE GOB-4 PROTEIN (ANTERIOR GRADIENT 2) (XENOPUS LAEVIS) (XENOPUS LAEVIS).
 GN AGR2 OR GOB-4 OR MAG-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE-INTESTINE;
 RA Komiyama T., Hirohashi S.;
 RT "Cloning of a gene, gob-4, which is expressed in intestinal goblet cells in mice.";
 RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN-SWISS WEBSTER;
 RX MEDLINE=99009231; PubMed=9790916;
 RA Thompson D.A., Weigel R.J.;
 RT "Mag-2, the human homologue of the Xenopus laevis cement gland gene xag-2, is coexpressed with estrogen receptor in breast cancer cell lines.";
 RT Blochem. Biophys. Res. Commun. 251:111-116(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP STRAIN-C57BL/6J; TISSUE-PANCREAS;
 RX MEDLINE=21085650; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Araiwa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I., Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schirml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seyer T., Shibata Y., Storch K.-F., Suzuki H., Togo-Oka K., Wang K.H., Weitz C., Wittaker C., Wilming L., Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RN Nature 409:685-690(2001).
 RP [4]
 RP SEQUENCE FROM N.A.

RC TISSUE-COLON, NORMAL. 5 MONTH OLD MALE MOUSE.
 RA Strassberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016592; BAA32044.1;
 DR EMBL; AF044262; AAC72705.1;
 DR EMBL; AC007677; BAB25181.1;
 DR EMBL; BC013334; AAH13334.1;
 DR MGD; MG1:134405; Agt2.
 SQ SEQUENCE 175 AA; 19920 MW; ACC3CFE429B668CA CRC64;

Query Match 77.9%; Score 828; DB 11; Length 175;
 Best Local Similarity 91.4%; Pred. No. 3.6e-73;
 Matches 160; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 32 MEKIPVSAPFLILVALSYTLARDTTPVPGAKKDKDSRPKLPQTLSRGWGQDLMTQTYEE 91
 1 MEKIPVSAPFLILVALSYTLARDTTPVPGAKKDKDSRPKLPQTLSRGWGQDLMTQTYEE 60

DB 92 ALYKSTSNKPLMIHHLDECPSHQAALKVFAENKEIQKLAEOQVFLNLVYETTDKHLSP 151
 61 ALYKSTSNKPLMIHHLDECPSHQAALKVFAENKEIQKLAEOQVFLNLVYETTDKHLSP 120

QY 152 DGOYVPRIMEFVPSLTVRADITGRYSNRLYAYEPADTALLDMKKALKLKTTEL 206
 121 DGOYVPRIMEFVPSLTVRADITGRYSNRLYAYEPADTALLDMKKALKLKTTEL 175

RESULT 3
 ID 090Y05 PRELIMINARY; PRT; 159 AA.
 AC 090Y05;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DE CEMENT GLAND-SPECIFIC PROTEIN CGS.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP Alexandrova E.M., Novoselov V.V., Zaratsky A.G.;
 RT "Three novel genes expressed in the anterior part of the Xenopus laevis embryo.";
 RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF314056; AAL26844.1;
 SQ SEQUENCE 159 AA; 18275 MW; 738D62284838B8EB CRC64;

Query Match 51.6%; Score 548; DB 13; Length 159;
 Best Local Similarity 60.0%; Pred. No. 8e-46;
 Matches 105; Conservative 26; Mismatches 28; Indels 16; Gaps 1;

QY 32 MEKIPVSAPFLILVALSYTLARDTTPVPGAKKDKDSRPKLPQTLSRGWGQDLMTQTYEE 91
 1 MEKIPVSAPFLILVALSYTLARDTTPVPGAKKDKDSRPKLPQTLSRGWGQDLMTQTYEE 44

QY 92 ALYKSTSNKPLMIHHLDECPSHQAALKVFAENKEIQKLAEOQVFLNLVYETTDKHLSP 151
 45 GLFKAKEENKPLMIHHLDECPSHQAALKVFAENKEIQKLAEOQVFLNLVYETTDKHLSP 104

DB 152 DGOYVPRIMEFVPSLTVRADITGRYSNRLYAYEPADTALLDMKKALKLKTTEL 206
 105 DGOYVPRIMEFVPSLTVRADITGRYSNRLYAYEPADTALLDMKKALKLKTTEL 159

RESULT 4
 ID 042251 PRELIMINARY; PRT; 185 AA.
 AC 042251;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)

DE CEMENT GLAND-SPECIFIC.
 GN XAG-2.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8335;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HEAD;
 RX MEDLINE=89324051; PubMed=2752418;
 RA Sive H.L., Hattori K., Weintraub H.;
 RT "Progressive determination during formation of the anteroposterior
 axis in Xenopus laevis.";
 RL Cell 58:171-180(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HEAD;
 RA Sive H.L., Mainstock D.H., Kennedy B.S.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF025474; AAB81968.1; -
 SO SEQUENCE 185 AA; 20498 MW; D6CDE02DEC38857B CRC64;

Query Match 42.1%; Score 447; DB 13; Length 185;
 Best Local Similarity 50.6%; Pred. No. 7.5e-36;
 Matches 89; Conservative 26; Mismatches 49; Indels 12; Gaps 2;

OY 43 LVALSYTARDTTVKRP-----GAKKDTKDSRPRLPOTLSRGWDQILMTQTYEE 91
 DB 10 LVLLSVLGEALRPRKROGATDTNGAKAKSEPAPKVTGLTDLRGCEDEIEMAQTYEE 69
 OY 92 ALYKSKTSKRLMTHHDECPHSQALKVFAENKEIOKLA-EQFLLMLVYETTDKHS 150
 DB 70 PGAKRNNKRLPLMTHHDECPSTALKKAFVADKMAQKLAODEFTMLNVHPVADENOS 129
 OY 151 PGQGVPRIMEFVPSLTIVRADITGRYSNRLVYEPADTLLLDNMKALKLKTTEL 206
 DB 130 PGHVPVRIFIDPSLTIVRSDLKGRGNKLVYADDDIELITNMKAKSFLKTTEL 185

RESULT 5
 O96H50 PRELIMINARY; PRT; 172 AA.
 AC O96H50;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DE HYPOTHEICAL 19.2 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=OVARY, AND ADENOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC008913; AAH08913.1; -
 KW Hypothetical protein.
 SO SEQUENCE 172 AA; 19228 MW; 5C9DDA6D5A7C4BDF CRC64;

Query Match 21.4%; Score 227; DB 4; Length 172;
 Best Local Similarity 34.8%; Pred. No. 2.3e-14;
 Matches 55; Conservative 25; Mismatches 36; Indels 22; Gaps 3;
 OY 66 DSRPKLPQT-----LSRGWDQILMTQTYEALYKSKTSNKPMLII 106
 DB 2 ETRPRIGATCLGFSFLLVISDGHNGKGGGHIHM-RTLEDGKKEAASGLPLMT 60
 OY 107 HHLDEPHSQALKVFAENKEIOKLAEOFYLLNLVYETTDKH--ISPDGOYVPRIMEFV 164
 DB 61 IHKSWGACKALKPKFAESTEISELSHNFVNVNLEDEEPRKHFDPDGGIPIRIILDP 120

OY 165 SLTVRADITGRYSNRLVYEPADTALLLDNMKALKLIL 202
 DB 121 SKGVHEITINENGNSYKFFYSABGVYGMKEAQERL 158

RESULT 6
 O9C0U0 PRELIMINARY; PRT; 170 AA.
 AC O9C0U0;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 0610040B21RIK PROTEIN (RIKEN CDNA 0610040B21 GENE).
 GN 0610040B21RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRIO, AND KIDNEY;
 RX MEDLINE=21083660; PubMed=11217851;
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okidell D., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK003481; BAB22811.1; -
 DR EMBL: AK002862; BAB22413.1; -
 DR EMBL: BC006857; AAH06857.1; -
 DR MCD: MGI:1913323; 0610040B21RIK.
 DR InterPro: IPR000886; ER-target.
 DR InterPro: IPR00063; Thioled.
 DR PROSITE: PS00014; ER-TARGET; UNKNOWN_1.
 SO SEQUENCE 170 AA; 19048 MW; 5B91FC9BE12C5E44 CRC64;

Query Match 21.1%; Score 224.5; DB 11; Length 170;
 Best Local Similarity 34.9%; Pred. No. 4e-14; 58; Indels 25; Gaps 4;
 Matches 58; Conservative 25; Mismatches 36; Indels 25; Gaps 4;

OY 39 AFLVVALSYTARDTTVKPKKDTKDSRPRLPOTLSRGWDQILMTQTYEALYKSKT 98
 DB 14 SFLLDITSS-----DGR-----TGLGFGFGHIHM-RTLEDGKKEAANA 50
 OY 99 SNKPLMTIHHDECPHSQALKVFAENKEIOKLAEOFYLLNL--VYETTDKHLSPDGQYV 156
 DB 51 SGLPLMTIHHKSGACKALKPKFAESTEISELSHNFVNVNLEDEEPRKHFDPDGQYI 110
 OY 157 PRIMEFVPSLTIVRADITGRYSNRLVYEPADTALLLDNMKALKLIL 202
 DB 111 PRILFDPGKVRPEITINENGNSYKFFYSABGVYGMKEAQERL 156

RESULT 7

```

095881 ID 095881 PRELIMINARY: PRT: 172 AA.
AC 095881;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE UNKNOWN (HYPOThETICAL 19.2 KDA PROTEIN).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96207227; PubMed=8619474;
RA Anderson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
RT "A 'double adaptor' method for improved shotgun library
RT construction."
RL Anal. Biochem. 236:107-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97264341; PubMed=9110174;
RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
RT "Large-scale concatenation cDNA sequencing."
RT Genome Res. 7:353-358(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX Mel G., Yu W., Gibbs R.A.;
RA Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX TISSUE-KIDNEY ADENOCARCINOMA;
RA Strausberg R.;
RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX TISSUE-COLON ADENOCARCINOMA;
RA Strausberg R.;
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF131758; AAD20035.1; -
DR EMBL: BC001493; AAH01493.1; -
DR EMBL: BC008953; AAH08953.1; -
DR InterPro: IPR000063; ThioRed.
DR PROSITE: PS00194; THIOREDOXIN; UNKNOWN_1.
KM Hypothetical protein.
SO SEQUENCE 172 AA; 19206 MW; 3092E9515A7C4094 CRC64;

Query Match 20.9%; Score 222; DB 4; Length 172;
Best Local Similarity 34.2%; Pred. No. 7,1e-14;
Matches 54; Conservative 26; Mismatches 56; Indels 22; Gaps 3;

OY 66 DSRKLPOT-----LSRGMDLIMTQTYEALYKSKTSNKKPLMI 106
DB 2 ETRRLATCLIGRSFLLIYSSDGHNGLGKGFEDHIHW-RTLEDGKKAASGLPLMTI 60
OY 107 HILDECPHSQALKKVFAPENKEIOKLAEOVLNLV--YETTDKHLSPDGQVPRIMEVDP 164
DB 61 IHSKWCAGKALKPKFAPSETEISLSHNFVWNLDEDEEPKDEDFSPDGQYIPRLFDLP 120
OY 165 SLTVRADITGRYSNRLAYEPADPALLLDNKKKALKL 202
DB 121 SGKVPHEIINENGMPYKYFYVSAOVVQGMKEAOERL 158

RESULT 8
O9NA78 ID 09NA78 PRELIMINARY: PRT: 186 AA.
AC 09NA78;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Y57A10A.23 PROTEIN.
GN Y57A10A.23.

```

```

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RT Science 282:2012-2018(1998).
DR EMBL: AL117195; CAB55026.1; -
DR InterPro: IPR000063; ThioRed.
SO SEQUENCE 186 AA; 21467 MW; F94E13D2F773321E CRC64;

Query Match 18.7%; Score 199; DB 5; Length 186;
Best Local Similarity 30.0%; Pred. No. 1.4e-11;
Matches 51; Conservative 38; Mismatches 57; Indels 24; Gaps 6;

OY 41 LLIVALSITLARDTYKPGAKKDKDSRPKLPOTLSRGMDLIMTQTYEALYKSKTSN 100
DB 4 LLLDAL-----VSASAYASFDKIKDSIQNPPLARGFDLDLAWYK-WEDALETALDID 53
OY 101 KPLMIHILDECPHSQALKKVFAPEN--KEIOKLAEOVLNLV--YETTDKHLSPDGQY 155
DB 54 KPIFLIHSKWCAGKALKPKFQOSNAKKAFFKLSHFVWNTEDDEPEEERYRPGKY 113
OY 156 VPRIMEVDPSTYRADITGRYSNRL-----YAYEPADTALLDNKKALK 200
DB 114 IPRLLFLDKN---GDILQEFKKKAEYKNYAYYSSPADILNSMKDVKL 159

RESULT 9
O9NS57 ID 09NS57 PRELIMINARY: PRT: 257 AA.
AC 09NS57;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOThETICAL 30.2 KDA PROTEIN.
GN F49H12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RT Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA Wohldmann P., Nhan M., Chisoe S., Hawkins J.;
RT "The sequence of C. elegans cosmid F49H12."
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RT EMBL: AC006642; AAF39830.1; -
DR InterPro: IPR000086; ER-target.
DR InterPro: IPR000063; ThioRed.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.

```


DB 310 CHIDPNCGLPR-PWMPWVEGAT---RPAVDVAYEPEVAERMTGALLNGMTR 360

RESULT 13

ID Q926X3 PRELIMINARY; PRT; 348 AA.

AC Q926X3;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PREDICTED DISULFIDE BOND ISOMERASE.
GN CPN0933 OR CPJ0933 OR CP0928.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=83558;

RP SEQUENCE FROM N.A.
RC STRAIN-CWL029;
RX MEDLINE=9206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Iamuel C., Fan J., Hyman R.W.,
RA Oliner L., Grimwood J., Davis R.W., Stephens R.S.,
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).

[2]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Ulfersack T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Linder M., Nelson W., Debo R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).

[3]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shihra M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CML029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).

DR EMBL: AE001673; AAD19071.1; -;
DR EMBL: AP002251; AAF38711.1; -;
DR EMBL: AP002548; BAA99141.1; -;
DR PHCT-2DPAGE; Q926X3; -;
DR TIGR; CP0928; -;
DR InterPro: IPR000063; Thioled.
KM Isomerase; Complete proteome.
SQ SEQUENCE 348 AA; 40492 MW; A7E03C043D7755C64;

Query Match 8.2%; Score 87; DB 16; Length 348;

Best Local Similarity 25.0%; Pred. No. 3;

Matches 29; Conservative 26; Mismatches 37; Indels 24; Gaps 4;

QY 29 RVAMEKIPVSAFLVLAATLADTVYKPKAKDTSRKLQTLRSRGDGLIWTQT 88

DB 9 RCSIKOKVLAATL--LISL--PTLEAENRDS-----DSIWHLD 46

QY 89 YEEALYKSKSNKPLMIHHLDEC--PHSQALKVFAENKEIOKLAQFVLLNLY 142

DB 47 YQELQSKKEALPLVIFSGSDWNGFCMKIRKVELSPETIKVQGRKVEVEY 102

RESULT 14

ID Q929J2 PRELIMINARY; PRT; 1151 AA.

AC Q929J2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE PEPTIDOLYCAN BOUND PROTEIN (LPXNG MOTIF).
GN LIN2283.

OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxId=1642;

[1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;

RA Glaeser P., Franjeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Domenech E., Dominguez-Bernal G., Duchaud E., Durand L., Dussurget O.,
RA Ertan K.-D., Faith H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Huf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapat G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;
RT "Comparative genomes of Listeria species.";
RL Science 294:849-852(2001).

DR EMBL: AL596171; CAC97511.1; -;
DR Listlist; LIN02283;
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 1151 AA; 127525 MW; 6A65870E5FBF6E1 CRC64;

Query Match 7.9%; Score 84.5; DB 16; Length 1151;

Best Local Similarity 22.8%; Pred. No. 26;

Matches 44; Conservative 31; Mismatches 51; Indels 67; Gaps 10;

QY 50 LARPTVYKPKAKK--DKKDSRKLPORTLSRGDGLIWTQ-----TYE 90

DB 961 VALDSVAPVTQYVDONHKQIASPELTGAGYGEKTPAKOKKITNTYLVKTPANVSTFN 1020

QY 91 EA-----LYKSKTSNKPMLIHHLDECPSQALKVFAENKEIOKLAQFVL--LNLV 141

DB 1021 EKATVTFVQKVTAGN--IIVDYD-----KNEKLAADSVILGKLNSS 1063

QY 142 YETTDKHLSPDQGY-VRPMEVPSLTVRADITGRYSNR---LYAEVPTALL----- 191

DB 1064 YRTSAKKISGYKLYQTPK-----NASCKFSNTSORVYVEKASTSISNKG 1111

QY 192 LDNMKKAALKIKT 204

DB 1112 ADNVEKSAKLPT 1124

RESULT 15

ID Q9BJK8 PRELIMINARY; PRT; 435 AA.

AC Q9BJK8;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NUCLEAR RECEPTOR NHR-79 (FRAGMENT).
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodieridae; Caenorhabditis.
OX NCBI_TaxId=6239;

[1]
RP SEQUENCE FROM N.A.

RA Bogan A., Maiba C.V., Yamamoto K., Cohen F., Sluder A.E.;
RT "Caenorhabditis elegans nuclear receptor sequences exhibit biophysical
RT compatibility with the ligand-binding domain fold.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: AF332208; AKI1979.1; -;
DR HSSP: P20393; 1A6T.
DR InterPro: IPR000536; Hormone_rec_119.

DR InterPro: IPRO01628; zf-C4.
DR Pfam: PF00104; hormone_rec. 1.
DR Pfam: PF00105; zf-C4; 1.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00399; zNF_C4; 1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW zinc-finger.
FT NOC_TFR 1 1
SQ SEQUENCE 435 AA; 50536 MW; B946B47CE5AE1492 CRC64;

Query Match	7.9%	Score 84;	DB 5;	Length 435;
Best Local Similarity	22.1%;	Pred. No. 7.8;		
Matches	25;	Conservative	26;	Mismatches 50;
				Indels 12;
				Gaps 3;

[illegible]

Search completed: October 9, 2002, 16:41:11
Job time : 29 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:39:22 : Search time 18 Seconds
(without alignments)
279.537 Million cell updates/sec

Title: US-09-674-266A-181

Perfect score: 1063

Sequence: 1 RLSCAGTUSGSGPHSRRLT.....DTALLDNMKKALKLTKTEL 206

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 100%
Listing first 45 summaries

Database :

- 1: Issued_Patents_AA:*
- 2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
- 3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
- 4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
- 5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
- 6: /cgn2_6/ptodata/2/1aa/6CTUS.COMB.pep:*
- 7: /cgn2_6/ptodata/2/1aa/6Cfills1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	899	84.6	175	4	US-08-916-576B-2
2	899	84.6	175	4	US-09-247-155-106
3	609.5	57.3	170	4	US-08-916-576B-8
4	605	56.9	131	4	US-09-247-155-174
5	555	52.2	166	4	US-08-916-576B-6
6	448	42.1	183	4	US-08-916-576B-7
7	222	20.9	172	4	US-08-916-576B-4
8	81.5	7.7	901	2	US-08-884-681-5
9	81.5	7.7	901	4	US-08-858-643-5
10	81.5	7.7	968	3	US-08-560-005-7
11	81.5	7.7	968	4	US-09-418-540-7
12	78	7.3	129	4	US-09-534-638-5
13	78	7.3	453	4	US-09-013-881-5
14	78	7.3	1160	5	PCR-US92-05401-4
15	77	7.2	652	2	US-08-313-185-53
16	77	7.2	652	2	US-08-459-499-17
17	77	7.2	652	2	US-09-082-614A-53
18	74	7.0	264	3	US-09-053-197A-8
19	74	7.0	264	4	US-09-085-761A-8
20	71.5	6.7	352	3	US-08-921-887-52
21	71	6.7	529	3	US-08-821-984-6
22	71	6.7	529	3	US-08-821-984-8
23	71	6.7	529	4	US-09-329-749-6
24	71	6.7	529	4	US-09-329-749-8
25	70.5	6.6	481	1	US-08-186-811-2
26	70.5	6.6	481	1	US-08-311-611A-98
27	70.5	6.6	481	1	US-08-372-783-98

28	70.5	6.6	481	1	US-08-372-105-98	Sequence 98, Appl
29	70.5	6.6	481	1	US-08-306-473A-98	Sequence 98, Appl
30	70.5	6.6	481	1	US-08-261-660A-4	Sequence 4, Appl
31	70.5	6.6	481	1	US-08-209-762-98	Sequence 98, Appl
32	70.5	6.6	481	1	US-08-473-344-98	Sequence 98, Appl
33	70.5	6.6	481	1	US-08-274-303-6	Sequence 6, Appl
34	70.5	6.6	481	1	US-08-377-391A-2	Sequence 2, Appl
35	70.5	6.6	481	2	US-08-485-445A-98	Sequence 98, Appl
36	70.5	6.6	481	2	US-08-779-400-2	Sequence 2, Appl
37	70.5	6.6	481	2	US-08-955-660-2	Sequence 2, Appl
38	70.5	6.6	481	3	US-09-119-263-98	Sequence 98, Appl
39	70.5	6.6	481	4	US-08-657-162-98	Sequence 98, Appl
40	70.5	6.6	481	4	US-09-224-480-98	Sequence 98, Appl
41	70.5	6.6	481	4	US-09-093-539-98	Sequence 98, Appl
42	70.5	6.6	481	4	US-09-146-620-2	Sequence 2, Appl
43	70.5	6.6	481	4	US-09-395-453-2	Sequence 2, Appl
44	70.5	6.6	481	5	PCR-US94-02465-98	Sequence 98, Appl
45	70.5	6.6	481	5	PCR-US94-06931-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-916-576B-2
Sequence 2, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNERS, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESSES:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentia Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916.576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024.347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 175 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-916-576B-2
Query Match 84.6%; Score 899; DB 4; Length 175;
Best Local Similarity 100.0%; Pred. NO. 8.5e-96;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 32 MEKIPASALLVALSYTLARDITVVRGAKKDTKDSRPLPTQLISGWCDOJLIMTQYEE 91

```

Db 1 MEXIPVSAFLLLVALSYTLARDTTPKGAKKDTPKDSRPKLPQTLSRGWGDLMTOTYEE 60
    |||
Qy 92 ALYKSTSNKPLMIITHLDECPSQALKVFAENKEIOKLAEOFVLNLVYETTDKHLSP 151
    |||
Db 61 ALYKSTSNKPLMIITHLDECPSQALKVFAENKEIOKLAEOFVLNLVYETTDKHLSP 120
    |||
Qy 152 DGQYVRIMEVDPSTLVRADITGRYSNRLYAYEPADTALLDNMKAALKLKTTEL 206
    |||
Db 121 DGQYVRIMEVDPSTLVRADITGRYSNRLYAYEPADTALLDNMKAALKLKTTEL 175
    |||

```

RESULT 2

```

US-09-247-155-106
; Sequence 106, Application US/09247155A*
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; EARLIER FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 106
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -20..-1
US-09-247-155-106

```

```

Query Match          84.6%; Score 899; DB 4; Length 175;
Best Local Similarity 100.0%; Pred. No. 8.5e-96;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 32 MEXIPVSAFLLLVALSYTLARDTTPKGAKKDTPKDSRPKLPQTLSRGWGDLMTOTYEE 91
    |||
Db 1 MEXIPVSAFLLLVALSYTLARDTTPKGAKKDTPKDSRPKLPQTLSRGWGDLMTOTYEE 60
    |||
Qy 92 ALYKSTSNKPLMIITHLDECPSQALKVFAENKEIOKLAEOFVLNLVYETTDKHLSP 151
    |||
Db 61 ALYKSTSNKPLMIITHLDECPSQALKVFAENKEIOKLAEOFVLNLVYETTDKHLSP 120
    |||
Qy 152 DGQYVRIMEVDPSTLVRADITGRYSNRLYAYEPADTALLDNMKAALKLKTTEL 206
    |||
Db 121 DGQYVRIMEVDPSTLVRADITGRYSNRLYAYEPADTALLDNMKAALKLKTTEL 175
    |||

```

RESULT 3

```

US-08-916-576B-8
; Sequence 8, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.

```

```

; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEFER, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 170 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-916-576B-8

```

```

Query Match          57.3%; Score 609.5; DB 4; Length 170;
Best Local Similarity 71.3%; Pred. No. 1.9e-62;
Matches 122; Conservative 14; Mismatches 26; Indels 9; Gaps 2;

```

```

Qy 37 VSAFLLLVALSYTLARDTTPKGAKKDTPKDSRP-KLPQTLSRGWGDLMTOTYEEALYK 95
    |||
Db 7 LGACLLVALSYXLA-----XXXKDXRPKGPQTLSRGWGDXKVVYEEGLK 58
    |||
Qy 96 SKTSNKPMLMIITHLDECPSQALKVFAENKEIOKLAEOFVLNLVYETTDKHLSPDGQY 155
    |||
Db 59 AXXSNKPLMIITHLDECPSQALKVFAENKEIOKLAEOFVLNLVYETTDENLSPDGQY 118
    |||
Qy 156 VPRIMEVDPSTLVRADITGRYSNRLYAYEPADTALLDNMKAALKLKTTEL 206
    |||
Db 119 VPRIMEVDPSTLVRADITGRYSNRLYAYEPDXDLIXNMKAALKLKTTEL 169
    |||

```

RESULT 4

```

US-09-247-155-174
; Sequence 174, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; EARLIER FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 174

```


LENGTH: 131
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: -20..-1
FEATURE:
NAME/KEY: UNSURE
LOCATION: 40,41,43,60,70,76,82,86,105,107
OTHER INFORMATION: Xaa = any one of the twenty amino acids
US-09-247-155-174

Query Match 56.9%; Score 605; DB 4; Length 131;
Best Local Similarity 92.9%; Pred. No. 4.2e-62;
Matches 117; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 32 MEKIPYSALLLVASTYLARDITYKPGAKKDTKSRPKLPOTLSRGWDQIMTQTYEE 91
DB 1 MEKIPYSALLLVASTYLARDITYKPGAKKDTKSRPKLPOTLSRGWDQIMTQTYEX 60
QY 92 ALYKSTSKRPMLTIHHDECPHSQALKKVFAPENKEIOKLAQFVLNLVYETTDKHLSP 151
DB 61 XLKSTSKRPMLTIHHDXCPHSQALKKXFAENKXIQKLAQFVXNLVYETTDKHLSP 120
QY 152 DQGYVP 157
DB 121 DQGYXP 126

RESULT 5

US-08-916-576B-6
Sequence 6, Application US/08916576B
Patent No. 6171816

GENERAL INFORMATION:

APPLICANT: YU, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 166 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-916-576B-6

Query Match 52.2%; Score 555; DB 4; Length 166;
Best Local Similarity 64.7%; Pred. No. 3.5e-56;
Matches 108; Conservative 28; Mismatches 19; Indels 12; Gaps 4;

QY 41 LLLVALSTYLARDITYKPGAKKDTKSRPKLPOTLSRGWDQIMTQTYEALYKSTISN 100
DB 11 LLLVYSNLA--IAIK-----KEKRP--POTLSRGWDQIMTQTYEGLFYAOKSK 59
QY 101 KPLMTHHDECPHSQALKKVFAPENKEIOKLA-QVLLNLVYETTDKRLSPDQGVPR 159
DB 60 KPLMTHHDECPHSQALKKVFAPENKEIOKLA-QVLLNLVYETTDKRLSPDQGVPR 119
QY 160 MEVDPSTLYRADITGRYSNRLVAYEPADTALLDNKKALKLITEL 206
DB 120 MEVDPSTLYRADITGRYSNRLVAYEPADTALLDNKKALKLITEL 166

RESULT 6

US-08-916-576B-7
Sequence 7, Application US/08916576B
Patent No. 6171816

GENERAL INFORMATION:

APPLICANT: YU, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 183 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-916-576B-7

Query Match 42.1%; Score 448; DB 4; Length 183;
Best Local Similarity 49.4%; Pred. No. 8.6e-44;
Matches 88; Conservative 29; Mismatches 51; Indels 10; Gaps 3;

QY 39 AFLIVLALSTYLARDITYVP---GAKKDTKSRPKLP-----QTLSRGWGQIMTQTY 89
DB 6 SLVGLVLLCSALGEAVLAKPKKQAGTTDRKTPAPIKTKGLKTLDRGWGESIEWQTY 65
QY 90 EALYKSTSKRPMLTIHHDECPHSQALKKVFAPENKEIOKLA-EQFVLNLVYETTDK 148

Db 66 EELAKARENKPLMVIHHEDECPYSIALKKAFAVADRAQKLAQEDFIMLVHPVADEN 125
QY 149 LSPDGVPRIMFVPSLTGRADITGRYSNRLAYEPADALLDNKKALKLTEL 206
Db 126 QSPDGHVPRVIFIDPSLTGRVSDLKGRYGNKMYADADDIPELITNKKAKSFLTEL 183

RESULT 7

US-08-916-576B-4
Sequence 4, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEEFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 172 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-916-576B-4

Query Match 20.9%; Score 222; DB 4; Length 172;
Best Local Similarity 34.2%; Pred. No. 8.8e-18;

Matches 54; Conservative 26; Mismatches 56; Indels 22; Gaps 3;

QY 66 DSRKLPQT-----LSRGMDQLMTQTYEALYKSTNKKPLMTI 106
Db 2 ETRRLGATCLGTFLLIVSSDGNGLGKGFQDHIW-RTLEDGKKAASGLPLMTI 60
QY 107 HHLECPHSAQLKVFENKEIOKLAQFVLLN--YETTDKHLSPDGVPRIMFVDP 164
Db 61 IHKSMCGACKALKPKFAESTIEISHNFVWVNLDEDEEPDEDFSPDGVPRILFLDP 120
QY 165 SLTVRADITGRYSNRLAYEPADTALLDNKKALKL 202
Db 121 SGKVHPELINENGNPSYKYFYVSAEOYVQGMKEAQR 158

RESULT 8

US-08-884-681-5
Sequence 5, Application US/08884681

Patent No. 5955338
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Cortley, Neil C.
TITLE OF INVENTION: NEW HUMAN PHOSPHATIDYLINOSITOL
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/884,681
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0334 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 901 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1420920
US-08-884-681-5

Query Match 7.7%; Score 81.5; DB 2; Length 901;
Best Local Similarity 23.7%; Pred. No. 1.6;

Matches 45; Conservative 32; Mismatches 46; Indels 67; Gaps 13;

QY 55 TVPGAKKDTKDSRPK--LPOTLSRGMDQLMTQTYEAL-YKS-----KTS-NKPLMTI 106
Db 476 TYKYDSKTDWSSGKCRVP-----AMCDRLIMRGTVNQLNRYSHMLKTSDBKPVSA 530
QY 107 HHL-----DECPHSAQLKVFENKEIOKLA-----QFVLNLYETTDK-- 147
Db 531 FHIGKYVDE--RRRKVFEDSVRIMDRMENFPLSLSRREYFENKKFQLOK 586
QY 148 -HLSPDG-----YVPRIM-----FVDSLTVRADITGRYSNRLAYE 184
Db 587 FQISNNGVPCVHFSTFKRLNDSQYCKPMLRAEPREGYLEPNEYV--DIS-----LDVV 638
QY 185 PADTALLDN 194
Db 639 SKDSVTLLNS 648

RESULT 9

US-09-258-643-5
Sequence 5, Application US/09258643
Patent No. 6277373
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Cortley, Neil C.


```
ZIP: 94105
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/418,540
: FILING DATE: 14-OCT-1999
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/560,005
: FILING DATE: 17-NOV-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Dow, Karen B.
: REGISTRATION NUMBER: 29,684
: REFERENCE/DOCKET NUMBER: 2307K-0624000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-326-2400
: TELEFAX: 415-326-2422
:
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 968 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FEATURE:
: NAME/KEY: Region
: LOCATION: 1..968
: OTHER INFORMATION: /note="ocr1"
:
: US-09-418-540-7

Query Match      7.7%; Score 81.5; DB 4; Length 968;
Best Local Similarity 23.7%; Pred. No. 1.8;
Matches 45; Conservative 33; Mismatches 46; Indels 67; Gaps 13;

OY 55 TVKGAKKDTRKSPK-LPOTLSRGWDQIWTQYEAL-YKS---KTS-NKPLMI 106
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 551 TYKXDSKTDNRWSSGKGVF-----AMCDRLMGTNVNQLNYSHMLKTSDBKPVSA 605
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 107 HHL-----DECPHQALKKVFAENKEIOKLAE-----QFVLNLVYETTK-- 147
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 606 FHIGKVVDE---RRRKVEEDSVRIKDMENDPLSLSESRREFEVENKFKQLOGK 661
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 148 -HLSPDGQ-----YVPRIM-----FVDPSLTVRADITGRYSNRLYAYE 184
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 662 FQIINNGVPCHEFSFIFRLNDSQYCKPWLRAPEEGYLEPNEYV--DIS-----LDVYV 713
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 185 PADTALLDN 194
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 714 SKDSVTILNS 723

RESULT 12
: US-09-534-638-5
: Sequence 5, Application US/09534638
: Patent No. 6320038
: GENERAL INFORMATION:
: APPLICANT: Panula, Pertti A.J.
: APPLICANT: Brandt, Annika
: APPLICANT: Westerland, Johanna
: TITLE OF INVENTION: Promoter for Neuropeptide FF Promoter and use thereof
: FILE REFERENCE: 2530-104
: CURRENT APPLICATION NUMBER: US/09/534,638
: CURRENT FILING DATE: 2000-03-27
: EARLIER APPLICATION NUMBER: 09/365755
: EARLIER FILING DATE: 1999-08-03
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 5
: LENGTH: 129
: TYPE: PRT
```

```
: ORGANISM: Mouse
: US-09-534-638-5

Query Match      7.3%; Score 78; DB 4; Length 129;
Best Local Similarity 22.9%; Pred. No. 0.23;
Matches 16; Conservative 14; Mismatches 40; Indels 0; Gaps 0;

OY 44 VALSYTLARDITTVKGAKKDTRKSPKLPOTLSRGWDQIWTQYEALYKSKTSNKP 103
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 IALFYILLCSNTSPSSRNSRSGSPQAPVYVSDSDREDYVLAVALQKCLLSPSSQFSM 62
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 104 MIHHDECP 113
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 63 TPVSRLSQSP 72

RESULT 13
: US-09-013-881-5
: Sequence 5, Application US/09013881
: Patent No. 6132964
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Lal, Preeti
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Corley, Neil C.
: APPLICANT: Guegler, Karl J.
: APPLICANT: Shah, Purya
: TITLE OF INVENTION: HUMAN HYDROLASE-LIKE MOLECULES
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/013,881
: FILING DATE: HERewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: BILLINGS, LUCY J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0470 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
:
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 453 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: LUNGNOT10
: CLONE: 1376382
:
: US-09-013-881-5

Query Match      7.3%; Score 78; DB 4; Length 453;
Best Local Similarity 22.5%; Pred. No. 1.5;
Matches 45; Conservative 34; Mismatches 79; Indels 42; Gaps 9;

OY 25 VRKSRVAMEKIPVSAFLVALSTLTARDITTVKGAKKDTRKSPKLPOTLSRGWQOLI 84
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Db 187 LRKGFYSQKVTNPNLRIISLNTNLYYXPNIMLTKTDP-----ANOFE 231
QY 85 WFOFEALYKSKTSNKRLMIHH--LDECPHSQ---ALKKVEAF-----NKELOKLA 133
Db 232 W---LESTLNNSOOKREYVYIAHPVGYLPSSONITAMREYVNEKLIIDIPKYSDVLAG 288
QY 134 QF---VLLNLVYETTDKHLSPDQGVYPRIMEVPSLTVRADITGRYSN---RLYAEPA 186
Db 289 QFYGHTHSDLSIMVLSDKKSPVNS-----LFVAPAVTVPKVSLKEQNTNPGIRLFQYDPR 343
QY 187 DFALLDNMKKALKILKTEL 206
Db 344 DYK-LIDMLQYVNLMTFANL 362

RESULT 14
PCT-US92-05401-4
Sequence 4, Application PC/TUS9205401
GENERAL INFORMATION:
APPLICANT: Lemisccha, Ibor R.
TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMCLONE SYSTEMS INCORPORATED
STREET: 180 VARICK STREET
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10014

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/05401
FILING DATE: 19920626
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Felt, Irving N.
REGISTRATION NUMBER: 28,601
REFERENCE/DOCKET NUMBER: LEM-3-PPPPPT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-645-1405
TELEFAX: 212-645-2054
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1160 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-05401-4

Query Match 7.3%; Score 78; DB 5; Length 1160;
Best Local Similarity 28.4%; Pred. No. 5.8;
Matches 27; Conservative 17; Mismatches 35; Indels 16; Gaps 4;

QY 68 RPKLQTSRGGDQILMTQTYEALYKSKTSNKRLMIHHLDECPHSALKKVAENKE 127
Db 952 RRSFP-NITSLGCOL---ADAEEAMYN-----VDGRAVSECPHYONRRPFSREND 999

QY 128 IOKLAQFVL---LNLVYETTDKHLSPDQGVYPR 158
Db 1000 LGLLSPQAOVEDSRNNLVLRITSSLHLSTLGCRLPK 1034

RESULT 15
US-08-313-185-53
Sequence 53, Application US/08313185
Patent No. 5851763
GENERAL INFORMATION:
APPLICANT: Heym, Beate

APPLICANT: Cole, Stewart
APPLICANT: Young, Douglas
APPLICANT: Zhang, Ying
APPLICANT: Honore, Nadine
APPLICANT: Telenti, Amalio
APPLICANT: Bodmer, Thomas
TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance
TITLE OF INVENTION: In Mycobacterium Tuberculosis
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,185
FILING DATE: 12-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 02356.0068-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-313-185-53

Query Match 7.2%; Score 77; DB 2; Length 652;
Best Local Similarity 20.7%; Pred. No. 3.2;
Matches 49; Conservative 32; Mismatches 84; Indels 72; Gaps 11;

QY 12 GPHSRRLTQGRWVKKRYAMKIVSAPLLVALSTYLARDTYKPPAKKDTKRSRKL 71

Db 211 GPNHS-PLSAAAIPTFRGMNDEETVA--LIAGHTLGKTHGAGPASHGPPPEADIE 266

QY 72 POTSRLGMD-----OLIMTQT-----YEEALYKS----- 96

Db 267 AGL--GNASSYSGVGADATISGEVWVTQTPTQNNFENLFENVLTKSPAGQAEAVDG 324

QY 97 -----KTSNKPIMTIHHLDEC-PSQALKKVAENKEIQKLAQFVLLNLVYET 144

Db 325 APDIIPDPDPSEKRRKPMVLVTDLLRFDPREYKISRRLNDEPFAFARAW-----FKL 378

QY 145 TDKHLSPD---GQYVPR--IMVDPDSLVRADI-----TGRSNRLYAEPADT 188
Db 379 THRDGPKRYIGPEVAKEDLIWQDPPOYPTEDIILKAIAAGLVSSEIVASASAST 435

Search completed: October 9, 2002, 16:41:41
Job time : 19 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:40:27 ; Search time 32 Seconds
(without alignments)

715.038 Million cell updates/sec

Title: US-09-674-266a-181

Perfect score: 206
Sequence: 1 RLSGAGTSLSGSGPHPSRRLT.....DFALLDNMKALILKTEL 206

Scoring tables: Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq_032802:*

- 1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	206	100.0	206	20	AAV73837
2	194	94.2	194	20	AAV76590
3	194	94.2	202	22	AAV75621
4	175	85.0	175	19	AAW77365
5	175	85.0	175	19	AAW37844
6	175	85.0	175	19	AAW37872
7	175	85.0	175	20	AAV59675
8	175	85.0	175	21	AAW24070
9	175	85.0	175	22	AAU08804
10	175	85.0	175	22	AAW72203
11	126	61.2	132	21	AAW00109

12	116	56.3	116	20	AAV12312
13	115	55.8	115	21	AAV64672
14	59	28.6	131	20	AAV59718
15	55	26.7	56	20	AAV11881
16	42	20.4	62	20	AAV11882
17	29	14.1	89	20	AAV76568
18	24	11.7	166	19	AAW37846
19	24	11.7	166	21	AAW00194
20	24	11.7	166	22	AAU25728
21	24	11.7	166	22	AAW72205
22	24	11.7	166	22	AAW31192
23	24	11.7	166	22	AAU07647
24	24	11.7	168	22	AAW24502
25	17	8.3	56	20	AAV11939
26	14	6.8	39	21	AAW58921
27	11	5.3	11	22	AAU08805
28	11	5.3	11	22	AAU08806
29	11	5.3	11	22	AAU25804
30	11	5.3	11	22	AAU25805
31	10	4.9	11	22	AAU25806
32	10	4.9	70	20	AAV11654
33	10	4.9	180	22	AAV73737
34	9	4.4	9	22	AAU25812
35	9	4.4	9	22	AAU07648
36	8	3.9	315	22	AAV73015
37	8	3.9	1284	22	ABW61690
38	7	3.4	10	22	AAW42189
39	7	3.4	73	22	AAW86554
40	7	3.4	105	22	AAW91788
41	7	3.4	105	22	AAO04178
42	7	3.4	109	21	AAW03933
43	7	3.4	109	22	AAW40426
44	7	3.4	109	22	AAW80983
45	7	3.4	119	22	AAW25807

ALIGNMENTS

RESULT 1
AAV73837
ID AAV73837 standard; Protein: 206 AA.
XX
AC AAV73837;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor EST fragment derived protein #24.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
treatment.
XX
OS Homo sapiens.
XX
PN DE19820190-A1;
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt C, Pilarsky C, Dahl E;
XX
DR WPI: 1999-621386/54.
XX
DR N-PSDB; AA252865.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
XX
PS proteins -
PS Claim 23; Page 318; 502pp; German.

XX This invention describes novel polypeptides and their encoding nucleic
CC acids derived from human pancreatic tumor tissue which have cytostatic
CC activity. The sequences are also useful in producing pharmaceutical
CC compositions for treatment of pancreatic tumors. AAY73814-174252
CC represent protein fragments encoded by the human pancreatic tumor cDNA
CC library derived expressed sequence tag (EST) sequences represented in
CC AA252858-253014.
XX
SQ Sequence 206 AA:
Query Match 100.0%; Score 206; DB 20; Length 206;
Best Local Similarity 100.0%; Pred. No. 5e-194;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RLSCAGTSSGSPHPSRRLLQGRVWKRKSRVAMEKIPVSAPLLVALSTYLARDTTVAPGA 60
DB 1 RLSCAGTSSGSPHPSRRLLQGRVWKRKSRVAMEKIPVSAPLLVALSTYLARDTTVAPGA 60
QY 61 KKDPRKSRPKLPOTLSRGWGDQIMTQTYEALYKSKTSNKPMLIIHHLDECPSQALKK 120
DB 61 KKDPRKSRPKLPOTLSRGWGDQIMTQTYEALYKSKTSNKPMLIIHHLDECPSQALKK 120
QY 121 VFAENKEIQKLAEGFVLLNLYETTDKHLSPDQYVPRIMEVDPSTLVRADITGRYSNRL 180
DB 121 VFAENKEIQKLAEGFVLLNLYETTDKHLSPDQYVPRIMEVDPSTLVRADITGRYSNRL 180
QY 181 YAYEPADTALLDNMKKALKLKTTEL 206
DB 181 YAYEPADTALLDNMKKALKLKTTEL 206
RESULT 2
AAY76590
ID AAY76590 standard; Protein: 194 AA.
XX
AC AAY76590;
XX
DT 10-APR-2000 (first entry)
XX
DE Human ovarian tumor EST fragment encoded protein 86.
XX
KW Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW gene therapy; treatment.
XX
OS Homo sapiens.
XX
XX DEL19817557-A1.
XX
XX 21-OCT-1999.
XX
XX 09-APR-1998; 98DE-1017557.
XX
XX 09-APR-1998; 98DE-1017557.
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX Rosenthal A, Specht T, Hinemann B, Schmitt A, Pilarsky C, Dahl E;
XX WPI; 1999-591920/51.
XX
XX N-PSDB; AA277486.
XX
XX New nucleic acid sequences expressed in ovarian, and some other, cancer
PT tissues, and derived polypeptides, for treatment of ovarian cancer and
PT identification of therapeutic agents -
XX
XX Claim 25; Page 279; 310pp; German.
XX
XX This invention describes novel nucleic acid (cDNA) sequences (A) which
CC have anticancer activity and are highly expressed in ovarian tumor
CC tissue (and some also in testis and breast cancer tissue). The products
CC of the invention can be used for gene therapy. (A) are used (1) for
CC recombinant expression of polypeptides (B) and (11) to isolate complete

CC genes. (B) are used (1) to identify agents suitable for treatment of
CC ovarian cancer; (11) directly for treating this form of cancer
CC (including expression from gene therapy vectors) and (111) for generation
CC of specific antibodies. (A) are identified by assembling ESTs (expressed
CC sequence tags) from a particular tissue type before comparison of
CC expression patterns. This allows a significantly longer fragment of the
CC gene to be revealed, so should reduce the number of failures associated
CC with the fact that ESTs from different libraries may represent different
CC parts of the same unknown gene, distorting the estimated frequency of
CC occurrence in a particular tissue. AAY76505-Y76638 represent protein
CC fragments encoded by the human ovarian tumor cDNA library derived EST
CC fragments represented in AA277450-277572.
XX
SQ Sequence 194 AA:
Query Match 94.28%; Score 194; DB 20; Length 194;
Best Local Similarity 100.0%; Pred. No. 2.9e-182;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 PHSRRRLQGRVWKRKSRVAMEKIPVSAPLLVALSTYLARDTTVPAKKDTRSRPKLP 72
DB 1 PHSRRRLQGRVWKRKSRVAMEKIPVSAPLLVALSTYLARDTTVPAKKDTRSRPKLP 60
QY 73 QTLSSRGWGDQIMTQTYEALYKSKTSNKPMLIIHHLDECPSQALKKVAENKEIQKLA 132
DB 61 QTLSSRGWGDQIMTQTYEALYKSKTSNKPMLIIHHLDECPSQALKKVAENKEIQKLA 120
QY 133 EQFVLLNLYETTDKHLSPDQYVPRIMEVDPSTLVRADITGRYSNRLYAYEPADTALL 192
DB 121 EQFVLLNLYETTDKHLSPDQYVPRIMEVDPSTLVRADITGRYSNRLYAYEPADTALL 180
QY 193 DNMKKALKLKTTEL 206
DB 181 DNMKKALKLKTTEL 194
RESULT 3
AAG75621
ID AAG75621 standard; Protein: 202 AA.
XX
AC AAG75621;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6385.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
XX Homo sapiens.
XX
XX WO200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000MO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
XX
XX 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
XX
XX N-PSDB; AAH35026.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 7865-7866; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing of P.
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAH37789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC Seq ID NO:1027 to 1052, 7921 and 7922.
 CC
 SQ Sequence 202 AA;

Query Match 94.2%; Score 194; DB 22; Length 202;
 Best Local Similarity 100.0%; Pred. No. 3e-182;
 Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 PAPSRLTQGRVWRKSRVAMEKIPVSAPFLVLAALSTYLAARDTTVPKAKKDTKDSRPKL 72
 DB 9 PAPSRLTQGRVWRKSRVAMEKIPVSAPFLVLAALSTYLAARDTTVPKAKKDTKDSRPKL 68
 OY 73 QFLSRGWDQLIWTOYTEBALYKSKTSNKPMLIIHLLDECPSHQAALKVFAENKEIQKLA 132
 DB 69 QFLSRGWDQLIWTOYTEBALYKSKTSNKPMLIIHLLDECPSHQAALKVFAENKEIQKLA 128
 OY 133 EGFVLNLVYETTDKRLSDGQYVPRIMFVDPSTLYRAITGRYSNRLYAEPAADTALL 192
 DB 129 EGFVLNLVYETTDKRLSDGQYVPRIMFVDPSTLYRAITGRYSNRLYAEPAADTALL 188
 OY 193 DNKKKALKLKTLEL 206
 DB 189 DNKKKALKLKTLEL 202

RESULT 4
 AAM77365
 ID AAM77365 standard; Protein: 175 AA.

AC AAM77365;
 XX
 DT 21-DEC-1998 (first entry)
 XX
 DE zsig10 polypeptide.
 XX
 KW Human; mucous-mediated function; adhesion; tumour metastasis;
 KW bacterial colonisation; microbial infection; AIDS; cystic fibrosis;
 KW chronic obstructive pulmonary disease; asthma; Crohn's disease;
 KW sinonasal inflammatory disease; inflammatory bowel disease; bronchitis.
 XX
 OS Homo sapiens.
 XX
 PN WO9841627-A1.
 XX
 PD 24-SEP-1998.
 XX
 PF 18-MAR-1998; 98WO-US05251.
 XX
 PR 19-MAR-1997; 97US-0039631.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Sheppard PO;
 XX
 DR WPI; 1998-531566/45.
 DR N-PSDB; AAV59320.
 DR

XX
 PT New isolated mucous-associated polypeptide, zsig10 - used to develop
 PT products for treating e.g. tumour metastasis, microbial infections,
 PT cystic fibrosis, asthma, bronchitis or inflammatory bowel disease
 PS
 PS Claim 1; Page 82; 109pp; English.

CC The human polypeptide zsig10 is involved in mucous-mediated functions
 CC such as adhesion. The products of the invention can be used in the study
 CC and treatment of e.g. tumour metastasis, bacterial colonisation,
 CC susceptibility to and persistence of infection, microbial infections,
 CC AIDS, cystic fibrosis, chronic obstructive pulmonary disease, asthma,
 CC sinonasal inflammatory disease, inflammatory bowel disease, bronchitis,
 CC or Crohn's disease. The products can also be used for detection,
 CC diagnosis and drug screening.
 CC
 XX
 SQ Sequence 175 AA;

Query Match 85.0%; Score 175; DB 19; Length 175;
 Best Local Similarity 100.0%; Pred. No. 1.2e-163;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 MEKIPVSAPFLVLAALSTYLAARDTTVPKAKKDTKDSRPKLPTQLSRGWDQLIWTOYTE 91
 DB 1 MEKIPVSAPFLVLAALSTYLAARDTTVPKAKKDTKDSRPKLPTQLSRGWDQLIWTOYTE 60
 OY 92 ALYKSKTSNKPMLIIHLLDECPSHQAALKVFAENKEIQKLAEGFVLNLVYETTDKHLSP 151
 DB 61 ALYKSKTSNKPMLIIHLLDECPSHQAALKVFAENKEIQKLAEGFVLNLVYETTDKHLSP 120
 OY 152 DQGYVRIMFVDPSTLYRADITGRYSNRLYAEPAADTALLDNKKKALKLKTLEL 206
 DB 121 DQGYVRIMFVDPSTLYRADITGRYSNRLYAEPAADTALLDNKKKALKLKTLEL 175

RESULT 5
 AAM37844
 ID AAM37844 standard; Protein: 175 AA.

AC AAM37844;
 XX
 DT 28-AUG-1998 (first entry)
 XX
 DE Human XAG growth factor huxag-1.
 XX
 KW huxag-1; XAG; growth factor; colon cancer; tumour marker;
 KW breast disease; liver disease; lung disease; emphysema;
 KW wound healing; diagnosis; therapy; human.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT Peptide
 FT 1..20
 FT /label= Sig-peptide
 FT 21..175
 FT Protein
 FT /label= Mat.protein
 FT /note= "Claim 11"
 FT
 FT Peptide
 FT 20..43
 FT /label= Epitope
 FT /note= "Claim 12"
 FT 44..52
 FT Peptide
 FT /label= Epitope
 FT /note= "Claim 12"
 FT 61..72
 FT Peptide
 FT /label= Epitope
 FT /note= "Claim 12"
 FT 90..103
 FT Peptide
 FT /label= Epitope
 FT /note= "Claim 12"
 FT 113..125
 FT Peptide
 FT /label= Epitope
 FT /note= "Claim 12"
 FT 138..150
 FT Peptide

FT /label= Epitope
 FT /note= "Claim 12"
 XX
 PN WO9807749-A1.
 XX
 PD 26-FEB-1998.
 XX
 PF 22-AUG-1997; 97WO-US14139.
 XX
 PR 23-AUG-1996; 96WO-US13766.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Dillion PJ, Ebner R, Endress GA, Yu G;
 XX
 DR WPI; 1998-169093/15.
 DR N-PSDB; AAV19155.
 XX
 PT New isolated human XAG growth factor(s) - used to develop products
 PT for treating e.g. liver, lung or breast diseases or
 PT hyperproliferative disorders, e.g. cancer.
 XX
 PS Claim 1; Fig 1; 141pp; English.
 XX
 CC This polypeptide comprises huxAG-1, a member of a novel family of
 CC human growth factors also including huxAG-2 (see AAW37845) and
 CC huxAG-3 (see AAW37846). These proteins share homology with the
 CC XAG protein of Xenopus laevis, which is involved in embryogenesis
 CC and is expressed in adult tissue. huxAG-1 is specifically found
 CC in cancerous colon cells and may therefore be a growth factor for
 CC colon cancer. huxAG-1 cDNA (see AAV19155) was isolated from a cDNA
 CC library derived from human colon cancer tissue. Vectors, host
 CC cells, antibodies, and screening methods for identifying agonists
 CC and antagonists of huxAG-1 are provided. HuxAG polypeptides are
 CC growth factors and can be used to stimulate proliferation of cells.
 CC They can be used to stimulate the proliferation and differentiation
 CC of hepatocytes to alleviate or treat liver diseases and pathologies
 CC such as fulminant liver failure caused by cirrhosis, liver damage
 CC caused by viral hepatitis and toxic substances. They can also be
 CC used to stimulate or promote liver regeneration, e.g. after
 CC surgery. They can also be used to prevent and heal damage to the
 CC lungs caused by various pathological states. They can be used to
 CC stimulate proliferation and differentiation and promote the repair
 CC of alveoli and broncholar epithelium to prevent, attenuate, or
 CC treat acute or chronic lung damage, e.g. emphysema, which results
 CC in the progressive loss of alveoli, and inhalation injuries, e.g.
 CC resulting from smoke inhalation and burns, that cause necrosis of
 CC the broncholar epithelium and alveoli. They can also be used to
 CC stimulate the proliferation and differentiation of breast tissue
 CC and could therefore be used to promote healing of breast tissue
 CC injury due to surgery, trauma or cancer. Antagonists can be used
 CC to treat hyperproliferative disorders, including cancer, in
 CC particular hepatocellular carcinoma, osteosarcoma, breast cancer,
 CC or colon cancer. The products can also be used for detection and
 CC diagnosis.
 CC
 XX
 SQ Sequence 175 AA;
 XX
 QY Query Match 85.0%; Score 175; DB 19; Length 175;
 Db Best Local Similarity 100.0%; Pred. No. 1,2e-163; Indels 0; Gaps 0;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 32 MEKIPVSAFLLLVALSYTLARDTTVKKPGAKKDTKDSRPKLPQTLSRGWGDLIMTQYEE 91
 Db 1 MEKIPVSAFLLLVALSYTLARDTTVKKPGAKKDTKDSRPKLPQTLSRGWGDLIMTQYEE 60
 QY 92 ALYKSKTSNKKPLMTIIHILDECPSOALKKYFAENKEIOKLAEOQVLLNLVYETDKHLP 151
 Db 61 ALYKSKTSNKKPLMTIIHILDECPSOALKKYFAENKEIOKLAEOQVLLNLVYETDKHLP 120
 QY 152 DGOVPRIMFVDPSSLTVRADITGGRYSNRLYAYEPADTALLLDNNKKALKLKTTEL 206
 Db 121 DGOVPRIMFVDPSSLTVRADITGGRYSNRLYAYEPADTALLLDNNKKALKLKTTEL 175

RESULT 6
 AAW37872
 ID AAW37872 standard; Protein; 175 AA.
 XX
 AC AAW37872;
 XX
 DT 10-AUG-1998 (first entry)
 XX
 DE Human protein comprising secretory signal amino acid sequence 9.
 XX
 KW Human protein; secretory signal; nutritional source; cytokine;
 KW immunolytic; haematopoiesis; activin; inhibin; tumour; chemotactic;
 KW chemokinetic; thrombolytic; anti-inflammatory; inhibition;
 KW stomach cancer cell.
 XX
 OS Homo sapiens.
 XX
 XX
 PN WO9811217-A2.
 XX
 PD 19-MAR-1998.
 XX
 PF 12-SEP-1997; 97WO-JP03239.
 XX
 PR 13-SEP-1996; 96JP-0243060.
 XX
 PA (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
 XX
 DR WPI; 1998-207380/18.
 DR N-PSDB; AAV29047, AAV29048.
 XX
 PT Human proteins with secretory signal sequences - used to treat
 PT immune deficiencies, infections, tumours, and haematopoietic
 PT disorders, etc.
 XX
 PS Claim 1; Pages 79; 131pp; English.
 XX
 CC This is the amino acid sequence of a novel human protein comprising
 CC a secretory signal isolated from stomach cancer cells. Its proteins
 CC can be used as nutritional sources or supplements. The proteins may
 CC also have cytokine functions, immune modulating functions,
 CC haematopoiesis regulating activity, activin/inhibin regulating
 CC activity, chemotactic/chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity.
 CC
 XX
 SQ Sequence 175 AA;
 XX
 QY Query Match 85.0%; Score 175; DB 19; Length 175;
 Db Best Local Similarity 100.0%; Pred. No. 1,2e-163; Indels 0; Gaps 0;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 32 MEKIPVSAFLLLVALSYTLARDTTVKKPGAKKDTKDSRPKLPQTLSRGWGDLIMTQYEE 91
 Db 1 MEKIPVSAFLLLVALSYTLARDTTVKKPGAKKDTKDSRPKLPQTLSRGWGDLIMTQYEE 60
 QY 92 ALYKSKTSNKKPLMTIIHILDECPSOALKKYFAENKEIOKLAEOQVLLNLVYETDKHLP 151
 Db 61 ALYKSKTSNKKPLMTIIHILDECPSOALKKYFAENKEIOKLAEOQVLLNLVYETDKHLP 120
 QY 152 DGOVPRIMFVDPSSLTVRADITGGRYSNRLYAYEPADTALLLDNNKKALKLKTTEL 206
 Db 121 DGOVPRIMFVDPSSLTVRADITGGRYSNRLYAYEPADTALLLDNNKKALKLKTTEL 175
 RESULT 7
 AAY59675
 ID AAY59675 standard; Protein; 175 AA.

XX AC AAY59675;
 XX XX
 DT 18-JAN-2000 (first entry)
 XX DE Secreted protein 108-008-5-0-A6-FL.
 XX XX
 KM Secreted protein; fingerprint identification technique;
 KM chromosome mapping; human; hereditary disease; diagnosis; cancer;
 KM hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
 KM autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
 KM renal injury; amino aciduria; hypoglycaemia; male rat infertility;
 KM hypertension.
 XX XX
 OS Homo sapiens.
 XX XX
 PN WO940189-A2.
 XX XX
 PD 12-AUG-1999.
 XX XX
 PF 09-FEB-1999; 99WO-1B00282.
 XX XX
 PR 09-FEB-1998; 98US-0074121.
 PR 13-APR-1998; 98US-0081563.
 PR 10-AUG-1998; 98US-0096116.
 PR 04-SEP-1998; 98US-0099273.
 XX XX
 PA (GEST) GENSET.
 XX XX
 PI Bougueleret L, Duclet A, Dumas Milne Edwards J;
 XX XX
 DR WPI; 1999-600966/51.
 DR N-PSDB; AAZ40803.
 XX XX
 PT Extended cDNAs useful for expressing secreted proteins and to obtain
 PT specific antibodies -
 XX XX
 PS Claim 10; Page 199; 244pp; English.
 XX XX
 CC This sequence represents a human secreted protein of the invention.
 CC The extended cDNAs (or genomic DNAs obtainable from them) may be used to
 CC prepare PCR primers and probes. These are useful for forensic matching or
 CC positive identification by DNA sequencing. They may also be used in
 CC alternative fingerprint identification techniques. Antibodies against the
 CC proteins encoded by the extended cDNAs are useful in identification of
 CC tissue types or cell species, as well as identifying tissue specific
 CC soluble proteins. The sequences can be used for chromosome mapping and
 CC identification of genes associated with hereditary diseases or drug
 CC response. signal sequences from the cDNAs can be used in construction of
 CC secretion vectors. Other sequences derived from the extended cDNAs can be
 CC used to clone upstream genomic DNA sequences including promoters. This is
 CC in turn useful for identifying proteins that interact with promoter
 CC sequences. Some of the proteins may be useful in diagnosing and treating
 CC several disorders including, but not limited to: cancer, hyperlipidaemia,
 CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
 CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
 CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.
 CC XX
 SQ Sequence 175 AA;
 Query Match 85.0%; Score 175; DB 20; Length 175;
 Best Local Similarity 100.0%; Pred. No. 1.2e-163;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 DGQVPRIMEVPSLTVRADITGRYSNRLYAYEPADTALLDNKKALKLTTEL 175
 |||
 RESULT 8
 ID AAB24070
 ID AAB24070 standard; Protein; 175 AA.
 XX XX
 AC AAB24070;
 XX XX
 DT 29-JAN-2001 (first entry)
 XX XX
 DE Human PRO1030 protein sequence SEQ ID NO:40.
 XX XX
 KM Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
 KM proliferation; tumorigenesis; identification; cancer; cytostatic;
 KM neurotropic; neuroprotective; antiinflammatory; immunosuppressive;
 KM immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
 KM neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
 KM hypothalamic disorder; glandular disorder; macrophagal disorder;
 KM epithelial disorder; stromal disorder; blastocoeleic disorder;
 KM inflammatory disorder; immunologic disorder.
 XX XX
 OS Homo sapiens.
 XX XX
 PN WO200053755-A2.
 XX XX
 PD 14-SEP-2000.
 XX XX
 PF 06-JAN-2000; 2000WO-US00376.
 XX XX
 PR 08-MAR-1999; 99WO-US05028.
 PR 02-JUN-1999; 99WO-US12252.
 PR 23-JUN-1999; 99US-0141037.
 PR 07-JUL-1999; 99US-0143048.
 PR 26-JUL-1999; 99US-0145698.
 PR 30-NOV-1999; 99WO-US28313.
 PR 20-DEC-1999; 99WO-US30911.
 PR 05-JAN-2000; 2000WO-US00219.
 XX XX
 PA (GETH) GENENTECH INC.
 XX XX
 PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;
 PI Wetanabe CK, Wood WT;
 XX XX
 DR WPI; 2000-572270/53.
 DR N-PSDB; AAC58380.
 XX XX
 PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
 PT treatment, diagnosis and prevention of cancer -
 XX XX
 PS Claim 61; Fig 28; 286pp; English.
 XX XX
 CC The present invention describes an isolated antibody that binds to
 CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO355,
 CC PRO619, PRO717, PRO805, PRO830, PRO848, PRO943, PRO1005, PRO1009,
 CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
 CC PRO1287, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
 CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
 CC growth. The PRO polypeptides and nucleotides are useful in the
 CC treatment, diagnosis and prevention of cancer. The antibodies and other
 CC anti-tumour compounds may be used to treat various conditions, including
 CC those characterized by overexpression and/or activation of the amplified
 CC PRO genes. Exemplary conditions or disorders to be treated with such
 CC antibodies and other compounds include benign or malignant tumours
 CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
 CC colorectal, prostate, pancreatic, lung, ovula, thyroid, hepatic
 CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
 CC leukaemias and lymphoid malignancies, other disorders such as neuronal,
 CC glial, astrocytal, hypothalamic and other glandular, macrophagal,
 CC epithelial, stromal and blastocoeleic disorders, and inflammatory,
 CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
 CC primers and hybridisation probes used in the isolation of the human PRO

CC cancer. huxAG-1 and the identified antagonist are useful for treating
 CC cancer, in particular colon cancer. Detecting altered levels of huxAG-1
 CC and its polynucleotides are useful for diagnosing or detecting cancer in
 CC mammals. The gene encoding huxAG-1 is useful for monitoring human
 CC colorectal carcinoma. huxAG-1 nucleic acid molecules are also useful for
 CC chromosome identification. The present sequence represents the huxAG-1
 CC protein.

CC
 XX
 SQ Sequence 175 AA:

Query Match 85.0%; Score 175; DB 22; Length 175;
 Best Local Similarity 100.0%; Pred. No. 1.2e-163;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPVSAFLILVALSYTLARDTTPVKGAKKDTKDSRPKLPTLSRGWDQILMTQTYEE 91
 DB 1 MEKIPVSAFLILVALSYTLARDTTPVKGAKKDTKDSRPKLPTLSRGWDQILMTQTYEE 60
 QY 92 ALYKSTSNKPLMIITHHDECPHSQALKRVFAENKEIQKLAQFVLNLVYETTDKHLSP 151
 DB 61 ALYKSTSNKPLMIITHHDECPHSQALKRVFAENKEIQKLAQFVLNLVYETTDKHLSP 120
 QY 152 DQQYPRIMEFVDPISLTVRADITGRYSNRLYAYEPADTALLDNMKKALKLTEL 206
 DB 121 DQQYPRIMEFVDPISLTVRADITGRYSNRLYAYEPADTALLDNMKKALKLTEL 175

RESULT 11
 AAG00109

ID AAG00109 standard; Protein; 132 AA.

AC AAG00109;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 4190.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KM gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP103401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Mline Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR N-PSDB: AAC00115.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 13; SEQ ID 4190; 71pp + CD-ROM; English.

CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic

CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

CC
 XX
 SQ Sequence 132 AA:

Query Match 61.2%; Score 126; DB 21; Length 132;
 Best Local Similarity 100.0%; Pred. No. 1.2e-115;

Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPVSAFLILVALSYTLARDTTPVKGAKKDTKDSRPKLPTLSRGWDQILMTQTYEE 91
 DB 1 MEKIPVSAFLILVALSYTLARDTTPVKGAKKDTKDSRPKLPTLSRGWDQILMTQTYEE 60
 QY 92 ALYKSTSNKPLMIITHHDECPHSQALKRVFAENKEIQKLAQFVLNLVYETTDKHLSP 151
 DB 61 ALYKSTSNKPLMIITHHDECPHSQALKRVFAENKEIQKLAQFVLNLVYETTDKHLSP 120
 QY 152 DQQYYP 157
 DB 121 DQQYYP 126

RESULT 12

ID AAY12312 standard; Protein; 116 AA.

AC AAY12312;

DT 17-JUN-1999 (first entry)

DE Human 5' EST secreted protein SEQ ID NO:343.

DE Human; secreted protein; EST; expressed sequence tag; diagnosis;

KM forensic; gene therapy; chromosome mapping; signal peptide;

KM upstream regulatory sequence; cytokine activity; cell proliferation;

KM differentiation; haematopoiesis regulation; tissue growth regulation;

KM reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

KM thrombolytic; anti-inflammatory; tumour inhibition.

XX Homo sapiens.

XX MO9906548-A2.

XX 11-FEB-1999.

XX 31-JUL-1998; 98WO-1B01222.

XX 01-AUG-1997; 97US-0905135.

XX (GEST) GENSET.

XX Duclert A, Dumas Mline Edwards J, Lacroix B;

DR WPI: 1999-153778/13.

DR N-PSDB: AAX41145.

PT New nucleic acids encoding human secreted proteins - obtained from
 PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
 PT kidney, lung, umbilical cord, placenta and colon tissue
 XX
 PS Claim 27; Page 682; 824pp; English.

CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins, and encode the proteins given in AAY12261 to
 CC AAY12514, respectively. The proteins given represent the signal peptide
 CC and an N-terminal fragment of a secreted protein. The nucleic acid
 CC sequences can be used for producing secreted human gene products. They
 CC can also be used to develop products for diagnosis and therapy. The
 CC proteins obtained may have cytokine activity, cell
 CC proliferation/differentiation activity, haematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and

CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX
SQ Sequence 116 AA;

Query Match 56.3%; Score 116; DB 20; Length 116;
Best Local Similarity 100.0%; Pred. No. 6.7e-106; Indels 0; Gaps 0;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DQ 32 MEKIPVSAFLLLVALSTLARDTTPKPKAKKDTDSRPKLPQTLSRGNGDQLMTQTYEE 91
|||||
DQ 1 MEKIPVSAFLLLVALSTLARDTTPKPKAKKDTDSRPKLPQTLSRGNGDQLMTQTYEE 60
|||||

DQ 92 ALYKSKTSNKPMLTIHHLDECPHSQALKKVAENKEIQKLAEOFLVLLNLYETTDK 147
|||||
DQ 61 ALYKSKTSNKPMLTIHHLDECPHSQALKKVAENKEIQKLAEOFLVLLNLYETTDK 116
|||||

RESULT 13
AAV64672
ID AAV64672 standard; Protein: 115 AA.
XX
AC AAV64672;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST related polypeptide SEQ ID NO:833.
XX
XX Human: 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification.
XX
XX Homo sapiens.
XX OS
XX WO9953051-A2.
XX
XX 21-OCT-1999.
XX
XX 09-APR-1999; 99WO-IB00712.
XX
XX 09-APR-1998; 98US-0057719.
XX
XX 28-APR-1998; 98US-0069047.
XX
XX (GENSET).
XX PA
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-038446/03.
XX
XX N-PSDB; AA242286.
XX
XX Novel secreted protein 5' expressed sequence tag sequences used in
XX diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
XX Claim 3; Page 604; 837pp; English.

XX AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAV64651 to
CC AA243075 represent the EST-related proteins corresponding to AA242265 to
CC AA243075. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in

CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AA242249 to AA242264 and AAV64644 to AAV64650 represent
CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 115 AA;

Query Match 55.8%; Score 115; DB 21; Length 115;
Best Local Similarity 100.0%; Pred. No. 6.4e-105; Indels 0; Gaps 0;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DQ 32 MEKIPVSAFLLLVALSTLARDTTPKPKAKKDTDSRPKLPQTLSRGNGDQLMTQTYEE 91
|||||
DQ 1 MEKIPVSAFLLLVALSTLARDTTPKPKAKKDTDSRPKLPQTLSRGNGDQLMTQTYEE 60
|||||

DQ 92 ALYKSKTSNKPMLTIHHLDECPHSQALKKVAENKEIQKLAEOFLVLLNLYETTD 146
|||||
DQ 61 ALYKSKTSNKPMLTIHHLDECPHSQALKKVAENKEIQKLAEOFLVLLNLYETTD 115
|||||

RESULT 14
AAV59718
ID AAV59718 standard; Protein: 131 AA.
XX
XX AAV59718;
XX
XX 18-JAN-2000 (first entry)
XX
XX Secreted protein 78-21-1-B7-FL1.
XX
XX Secreted protein; fingerprint identification technique;
KW chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW hypertension.
XX
XX Homo sapiens.
XX OS
XX WO9940189-A2.
XX
XX 12-AUG-1999.
XX
XX 09-FEB-1999; 99WO-IB00282.
XX
XX 09-FEB-1998; 98US-0074121.
XX
XX 13-APR-1998; 98US-0081563.
XX
XX 10-AUG-1998; 98US-0096116.
XX
XX 04-SEP-1998; 98US-0099273.
XX
XX (GENSET).
XX PA
XX Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
XX WPI: 1999-600966/51.
XX
XX N-PSDB; AA240846.
XX
XX Extended cDNAs useful for expressing secreted proteins and to obtain
XX specific antibodies -
XX
XX Claim 10; Page 240; 244pp; English.

XX This sequence represents a human secreted protein of the invention.
CC The extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC prepare PCR primers and probes. These are useful for forensic matching or
CC positive identification by DNA sequencing. They may also be used in
CC alternative fingerprint identification techniques. Antibodies against the
CC proteins encoded by the extended cDNAs are useful in identification of
CC tissue types or cell species, as well as identifying tissue specific

CC soluble proteins. The sequences can be used for chromosome mapping and
 CC identification of genes associated with hereditary diseases or drug
 CC response. Signal sequences from the cDNAs can be used in construction of
 CC secretion vectors. Other sequences derived from the extended cDNAs can be
 CC used to clone upstream genomic DNA sequences including promoters. This is
 CC in turn useful for identifying proteins that interact with promoter
 CC sequences. Some of the proteins may be useful in diagnosing and treating
 CC several disorders including, but not limited to: cancer, hyperlipidaemia,
 CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
 CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
 CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.

CC Sequence 131 AA;

Query Match 28.6%; Score 59; DB 20; Length 131;
 Best Local Similarity 100.0%; Pred. No. 6.9e-50;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPVSAFLILVALSYTLARDTVKPGAKKDKSRPKLPOTLSRGWDQLIWT 90
 DB 1 MEKIPVSAFLILVALSYTLARDTVKPGAKKDKSRPKLPOTLSRGWDQLIWT 90

RESULT 15

AY11881
 ID AY11881 standard; Protein; 56 AA.

AC AY11881;

DT 18-JUN-1999 (first entry)

DE Human 5' EST secreted protein SEQ ID NO: 481.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.

PN MO9906550-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; 98WO-IB01232.

PR 01-AUG-1997; 97US-0905144.

PA (GEST) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR MPI: 1999-153780/13.

DR N-PSDB; AAX40603.

PT New isolated prostate-derived nucleic acids - used to develop
 PT products which may have cytokine, immune regulatory, haematopoiesis
 PT regulating, anti-inflammatory or tumour inhibition activity

PS Claim 34; Page 603; 675pp; English.

CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins expressed in prostate, and encode the proteins
 CC given in AY11716 to AY11993 respectively. The proteins given represent
 CC the signal peptide and an N-terminal fragment of a secreted protein. The
 CC nucleic acid sequences can be used for producing secreted human gene
 CC products. They can also be used to develop products for diagnosis and
 CC therapy. The proteins obtained may have cytokine activity, cell
 CC proliferation and differentiation activity, haematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and

CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity or other activities. The products
 CC can be used in forensic, gene therapy and chromosome mapping procedures.
 CC The sequences can also be used for obtaining corresponding promoter
 CC sequences. The nucleic acids encoding the signal peptides can be used for
 CC directing extracellular secretion of a polypeptide or the insertion of a
 CC polypeptide into a membrane, or importing a polypeptide into a cell.

CC Sequence 56 AA;

Query Match 26.7%; Score 55; DB 20; Length 56;
 Best Local Similarity 100.0%; Pred. No. 2.6e-46;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPVSAFLILVALSYTLARDTVKPGAKKDKSRPKLPOTLSRGWDQLIWT 86
 DB 1 MEKIPVSAFLILVALSYTLARDTVKPGAKKDKSRPKLPOTLSRGWDQLIWT 55

Search completed: October 9, 2002, 16:42:59
 Job time : 32 secs

A:Status: preliminary
A:Molecule type: protein
A:Residues: 144-151:270-277 <TM02>
A:Experimental source: taste bud
C:Comment: This protein is coupled to a GTP-binding protein-mediated signal transduction
C:Genetics:
A:Gene: tb567
C:Superfamily: olfactory receptor OR14
C:Keywords: olfaction; taste bud; transmembrane protein
F:26-49/Domain: transmembrane #status predicted <TM1>
F:58-79/Domain: transmembrane #status predicted <TM2>
F:101-120/Domain: transmembrane #status predicted <TM3>
F:140-164/Domain: transmembrane #status predicted <TM4>
F:197-219/Domain: transmembrane #status predicted <TM5>
F:238-260/Domain: transmembrane #status predicted <TM6>
F:273-293/Domain: transmembrane #status predicted <TM7>

Query Match 3.9%; Score 8; DB 2; Length 315;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LLLVALSY 48
Db 142 LLLVALSY 149

RESULT 3
C75578
probable acyl-CoA dehydrogenase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence-revision 03-Dec-1999 #text-change 31-Mar-2000
C:Accession: C75578
R:White, O.; Eisen, J.A.; Helgelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.S.; Shen, M.; Vamthavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zaleski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; M0ID:20036896
A:Accession: C75578
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1637 <WHI>
A:Cross-references: GB:AE001863; GB:AE001825; NID:96460670; PIDN:AAF12388.1; PID:9646068
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0250
A:Map position: 2

Query Match 3.9%; Score 8; DB 2; Length 637;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 186 ADTALLLD 193
Db 422 ADTALLLD 429

RESULT 4
C72486
hypothetical protein APE2531 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence-revision 20-Aug-1999 #text-change 09-Jun-2000
C:Accession: C72486
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; M0ID:99310339
A:Accession: C72486
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-101 <RAW>
A:Cross-references: DDBJ:AF000064; NID:95105945; PIDN:BAAB1547.1; PID:01045333; PID:9510

A:Experimental source: strain K1
C:Genetics:
A:Gene: APE2531
C:Superfamily: Aeropyrum pernix hypothetical protein APE2531

Query Match 3.4%; Score 7; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 AGTLGSG 11
Db 11 AGTLGSG 17

RESULT 5
XLMH
collipase precursor [validated] - human
N:Alternate names: procollipase
C:Species: Homo sapiens (man)
C>Date: 04-Dec-1986 #sequence-revision 19-May-1995 #text-change 08-Dec-2000
C:Accession: A42568; A33949; A03163
R:Sims, H.F.; Lowe, M.E.
Biochemistry 31, 7120-7125, 1992
A:Title: The human collipase gene: isolation, chromosomal location, and tissue-specific
A:Reference number: A42568; M0ID:92353041
A:Accession: A42568
A:Molecule type: DNA
A:Residues: 1-112 <SIM>
A:Cross-references: GB:M95529; NID:9180842; PIDN:AA05818.1; PID:91483624
A:Note: Sequence extracted from NCBI backbone (NCBIN:110576, NCBIN:110578, NCBIP:1105
R:Lowe, M.E.; Rosenblum, J.L.; McEwen, P.; Strauss, A.W.
Biochemistry 29, 823-828, 1990
A:Title: Cloning and characterization of the human collipase cDNA.
A:Reference number: A33949; M0ID:90248429
A:Accession: A33949
A:Molecule type: mRNA
A:Residues: 1-112 <LOW>
A:Cross-references: GB:J02883; NID:9180885; PIDN:AA52054.1; PID:9180886
A:Note: Evidence of partial N-glycosylation, possibly at Asn-43
R:Sternby, B.; Engstrom, A.; Hellman, U.; Vihert, A.M.; Sternby, N.H.; Borgstrom, B.
Biochim. Biophys. Acta 784, 75-80, 1984
A:Title: The primary sequence of human pancreatic collipase.
A:Reference number: A90652; M0ID:84104937
A:Accession: A03163
A:Molecule type: protein
A:Residues: 23-108 <STR>
C:Comment: Collipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms a 1:1 s
e the enzyme is washed off by bile salts, which are known to have an inhibitory effe
C:Genetics:
A:Gene: GDB:CLPS
A:Cross-references: GDB:127277; OMIM:120105
A:Map position: 6pter-6p21.1
A:Intons: 28/3; 69/3
C:Superfamily: collipase
C:Keywords: lipid digestion; lipid hydrolysis; pancreas
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-22/Domain: amino-terminal propeptide #status predicted <APP>
F:23-108/Product: collipase #status experimental <MAT>
F:109-112/Domain: carboxyl-terminal propeptide #status predicted <CPP>
F:34-104.40-56.44-80.45-78.66-86/Disulfide bonds: #status predicted
F:69,72,75,76/Binding site: micellar substrate (Lys, Tyr, Tyr, Tyr) #status predicted

Query Match 3.4%; Score 7; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LLLVALS 47
Db 7 LLLVALS 13

RESULT 6
B95024

conserved domain protein SP0207 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: B95024
R:Retellein, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95024
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-112 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74387.1; PID:q14971675; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0207

Query Match 3.4%; Score 7; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 50 LARDTIV 56
|||||||
DB 60 LARDTIV 66

RESULT 7
D90784
hypothetical protein ECs1244 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: D90784
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90784
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-114 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA834667.1; PID:q13360704; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs1244
C:Superfamily: hypothetical protein b3024

Query Match 3.4%; Score 7; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 KALKLTK 203
|||||||
DB 108 KALKLTK 114

RESULT 8
E90903
hypothetical protein ECs2197 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: E90903
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90903
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-114 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA835620.1; PID:q13361663; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs2197
C:Superfamily: hypothetical protein b3024

Query Match 3.4%; Score 7; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 KALKLTK 203
|||||||
DB 108 KALKLTK 114

RESULT 9
E85644
hypothetical protein Z1498 [imported] - Escherichia coli (strain O157:H7, substrain E
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85644
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85644
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-114 <STO>
A:Cross-references: GB:AE005174; NID:q12514358; PIDN:AAG55617.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDU933
C:Genetics:
A:Gene: Z1498
C:Superfamily: hypothetical protein b3024

Query Match 3.4%; Score 7; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 KALKLTK 203
|||||||
DB 108 KALKLTK 114

RESULT 10
E85713
unknown protein encoded within prophage CP-9330 [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85713
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85713
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-114 <STO>
A:Cross-references: GB:AE005174; NID:q12515051; PIDN:AAG56169.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDU933
C:Genetics:
A:Gene: Z2099
C:Superfamily: hypothetical protein b3024

Query Match 3.4%; Score 7; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 KALKLTK 203
|||||||

Db 108 KALKLKL 114

RESULT 11

H83514

conserved hypothetical protein PA1059 [Imported] - *Pseudomonas aeruginosa* (strain PA01)

C:Species: *Pseudomonas aeruginosa*

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Feb-2002

C:Accession: H83514

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mitsuuchi, S.D.; Warrenner, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

., Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho

A:Reference number: AB2950; MUID:20437337

A:Accession: H83514

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-115 <STO>

A:Cross-references: GB:AE004537; GB:AE004091; NID:g9946960; PIDN:AA604448.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1059

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1428

Query Match

Best Local Similarity 3.4%; Score 7; DB 2; Length 115;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 14 VSAFLLL 20

RESULT 12

P90432

hypothetical protein SS02594 [Imported] - *Sulfolobus solfataricus*

C:Species: *Sulfolobus solfataricus*

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: F90432

R:She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Ayanaz, M.J.; Chan

jong, I.; Jeffries, A.C.; Kosera, C.J.; Medina, N.; Peng, X.; Thi Ngoc, H.P.; Redder, F

arrault, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.

submitted to Genbank, April 2001

A:Description: *Sulfolobus solfataricus* complete genome.

A:Reference number: A99139

A:Accession: F90432

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-126 <KUR>

A:Cross-references: GB:AE006641; NID:g13815896; PIDN:AAK42717.1; GSPDB:GN00155

C:Genetics:

A:Gene: SS02594

Query Match

Best Local Similarity 3.4%; Score 7; DB 2; Length 126;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 85 VSAFLLL 91

RESULT 13

B41132

collagen-related protein 2 - *Hydra magnipapillata* (fragment)

C:Species: *Hydra magnipapillata*

C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 01-Dec-2000

C:Accession: B41132; S21930

R:Kurz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.

J. Cell Biol 115, 1159-1169, 1991

A:Title: Mini-collagens in hydra nematocytes.

A:Reference number: A41132; MUID:92064646

A:Accession: B41132

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-142 <KUR>

A:Cross-references: EMBL:X61046; NID:g9448; PIDN:CAA3380.1; PID:g9449

A:Note: submitted to the EMBL Data Library, July 1991

C:Superfamily: unassigned collagens

Query Match

Best Local Similarity 3.4%; Score 7; DB 2; Length 142;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 39 AFLLLVA 45

Db 2 AFLLLVA 8

RESULT 14

AH0652

conserved hypothetical protein STY1322 [Imported] - *Salmonella enterica* subsp. enteri

C:Species: *Salmonella enterica* subsp. enterica serovar Typh

A:Note: this species has also been called *Salmonella typhi*

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C:Accession: AH0652

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moule, S.; O'Garra, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se

A:Reference number: AB0502; PMID:11677608

A:Accession: AH0652

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-167 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08403.1; PID:g16502446; GSPDB:GN00176

C:Genetics:

A:Gene: STY1322

Query Match

Best Local Similarity 3.4%; Score 7; DB 2; Length 167;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 197 KALKLKL 203

Db 131 KALKLKL 137

RESULT 15

B97895

hypothetical protein spr0186 [Imported] - *Streptococcus pneumoniae* (strain R6)

C:Species: *Streptococcus pneumoniae*

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: B97895

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.

A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: B97895

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-205 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK98990.1; PID:g15457730; GSPDB:GN00174

C:Genetics:

A:Gene: spr0186

Query Match

Best Local Similarity 3.4%; Score 7; DB 2; Length 205;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 50 LARDTV 56
|||||
Db 153 LARDTV 159

Search completed: October 9, 2002, 16:44:40
Job time : 22 secs

FI	DISOLFID	80	104	BI	SIMILARITY.
SO	SEQUENCE	107	AA: 11271	MM: 825BA1AEB1422390	CRC64:

Query Match 3.4%; Score 7; DB 1; Length 107;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LLLVALS 47
 DB 7 LLLVALS 13

RESULT 2
 COL_HUMAN STANDARD; PRT; 112 AA.
 ID COL_HUMAN
 AC P04118;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Colipase precursor.
 GN CLPS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90248429; PubMed=2337598;
 RA Lowe M.E., Rosenblum J.L., McEwen P., Strauss A.W.;
 RT "Cloning and characterization of the human colipase cDNA."
 RL Biochemistry 29:823-828(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92353041; PubMed=1643046;
 RA Sims H.F., Lowe M.E.;
 RT "The human colipase gene: isolation, chromosomal location, and
 tissue-specific expression."
 RL Biochemistry 31:7120-7125(1992).
 RN [3]
 RP SEQUENCE OF 23-108.
 RX MEDLINE=84104937; PubMed=6691986;
 RA Sternby B., Engstrom A., Hellman U., Vihert A.M., Sternby N.-H.,
 RA Borgstrom B.;
 RT "The primary sequence of human pancreatic colipase."
 RL Biochim. Biophys. Acta 784:75-80(1984).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RX van Tilbeurgh H., Egloff M.-P., Martinez C., Rugani N., Verger R.,
 RA Cambilliau C.;
 RT "Interfacial activation of the lipase-procolipase complex by mixed
 micelles revealed by X-ray crystallography."
 RL Nature 362:814-820(1993).
 CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
 THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
 COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
 INHIBITORY EFFECT ON THE LIPASE.
 CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
 SIGNAL.
 CC -1- SUBUNIT: FORM A 1.1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; J02883; AAA52054.1; -;
 DR EMBL; M95529; AAB05818.1; -;
 DR PIR; A03163; XLHU.
 DR PIR; A33949; A33949.
 DR PIR; A42568; A42568.
 DR HSSP; P02703; 1ETH.
 DR MIM; 120105; -;

DR InterPro: IPR001981; Colipase.
 DR Pfam: PF01114; Colipase; 1.
 DR Pfam: PF02740; Colipase_C; 1.
 DR PRINTS: PR00128; COLIPASE.
 DR SMART: SM00023; COLIPASE; 1.
 DR PROSITE: PS00121; COLIPASE; 1.
 KW Lipid degradation; Digestion; Pancreas; Signal.
 FT SIGNAL 1 17
 FT PROPEP 18 22
 FT CHAIN 23 112
 FT DISULFID 34 45
 FT DISULFID 40 56
 FT DISULFID 44 78
 FT DISULFID 66 86
 FT DISULFID 80 104
 FT CONFLICT 68 69
 SQ SEQUENCE 112 AA; 11954 MW; 772872EBBE7C4DF8 CRC64;

Query Match 3.4%; Score 7; DB 1; Length 112;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LLLVALS 47
 DB 7 LLLVALS 13

RESULT 3
 XAG_XENLA STANDARD; PRT; 183 AA.
 ID XAG_XENLA
 AC P55868;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Putative secreted protein XAG precursor.
 GN XAG.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_Taxid=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Cement gland;
 RX MEDLINE=97003224; PubMed=8850563;
 RA Sive H., Bradley L.;
 RT "A sticky problem: the Xenopus cement gland as a paradigm for
 anteroposterior patterning."
 RL Dev. Dyn. 205:265-280(1996).
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN CEMENT GLAND.
 CC -1- SIMILARITY: HIGH, TO XENOPUS NP77.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; U76752; AAB18819.1; -;
 DR EMBL; U76752; AAB18819.1; -;
 FT SIGNAL 1 18
 FT CHAIN 19 183
 SQ SEQUENCE 183 AA; 20475 MW; 95E12B06FC8BDC8 CRC64;

POTENTIAL.
 PUTATIVE SECRETED PROTEIN XAG.

Query Match 3.4%; Score 7; DB 1; Length 183;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 DPSLTVR 169

DB 140 DPLSTVR 146

RESULT 4
NP77_XENLA STANDARD; PRT; 185 AA.

AC P55869;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Putative secreted protein NP77 precursor.
GN NP77.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cement gland;
RA Abberger F., Schueren C., Lepperdinger G., Richter K., Grunz H.;
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN CEMENT GLAND (POTENTIAL).
CC -1- SIMILARITY: HIGH TO XENOPUS XAG.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U82110; AAB49974.1; -
KW Signal.
FT CHAIN 1 185 PUTATIVE SECRETED PROTEIN NP77.
SQ SEQUENCE 185 AA; 20442 MW; AE3807C926044509 CRC64;

Query Match 3.4%; Score 7; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 DPLSTVR 169
DB 142 DPLSTVR 148

RESULT 5
ATTE_HYACE STANDARD; PRT; 235 AA.
AC P01513;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Attacin E and F precursor (Immune protein P5).
OS Hyalophora cecropia (Cecropia moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Saturniidae; Saturniinae; Hyalophora.
OX NCBI_TaxID=7123;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91160561; PubMed=2001705;
RA Sun S.C., Lindstrom I., Lee J.-Y., Faye I.;
RT "Structure and expression of the attacin genes in Hyalophora
RT cecropia.";
RL Eur. J. Biochem. 196;247-254(1991).
RN [2]
RP SEQUENCE OF 48-235 FROM N.A.

RA Kockum K., Faye I., von Hofsten P., Lee J.-Y., Xanthopoulos K.G.,
RA Boman H.G.;
RT "Insect Immunity. Isolation and sequence of two cDNA clones
RT corresponding to acidic and basic attacins from Hyalophora cecropia.";
RL EMBO J. 3:2071-2075(1984).
RN [3]

RP SEQUENCE OF 48-235 FROM N.A.
RX MEDLINE=86005745; PubMed=3840100;
RA Boman H.G., Faye I., von Hofsten P., Kockum K., Lee J.-Y.,
RA Xanthopoulos K.G., Bannich H., Engstroem A., Merrifield R.B.,
RA Andreu D.;
RT "On the primary structures of lysozyme, cecropins and attacins from
RT Hyalophora cecropia.";
RL Dev. Comp. Immunol. 9:551-558(1985).

RP SEQUENCE OF 48-231.
RA Engstroem A., Engstroem P., Tao Z.-J., Carlsson A., Bannich H.;
RT "Insect immunity. The primary structure of the antibacterial protein
RT attacin F and its relation to two native attacins from Hyalophora
RT cecropia.";
RL EMBO J. 3:2065-2070(1984).

CC -1- FUNCTION: ATTACINS ARE HEMOLYMPH ANTIBACTERIAL PROTEINS.
CC -1- PTM: ATTACIN F APPEARS TO BE DERIVED BY PROTEOLYTIC DIGESTION OF
CC ATTACIN E.
CC -1- MISCELLANEOUS: THERE ARE SIX FORMS OF ATTACIN THAT ARE DIVIDED
CC INTO TWO GROUPS: ACIDIC (E AND F) AND BASIC (A, B, C, AND D).
CC -1- SIMILARITY: BELONGS TO THE ATTACIN/SARCOTOXIN II FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X57715; CAA40886.1; -
DR EMBL; X00869; CAA25414.1; -
DR EMBL; M34926; AAA29183.1; -
DR PIR; A01775; EMMKRC.
DR PIR; S14104; S14104.

KW Insect immunity; Antibiotic; Hemolymph; Multigene family; Signal;
Repeat.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 47
FT CHAIN 48 235 ATTACIN E.
FT CHAIN 48 231 ATTACIN F.
FT DOMAIN 104 169 GLY-RICH (G1).
FT DOMAIN 170 235 GLY-RICH (G2).
SQ SEQUENCE 235 AA; 25437 MW; 35E1B549FE0417D1 CRC64;

Query Match 3.4%; Score 7; DB 1; Length 235;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 40 FULLVAL 46
DB 7 FULLVAL 13

RESULT 6
MCTL_SHEEP STANDARD; PRT; 245 AA.
ID P80931;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell protease 1A precursor (BC 3.4.21.-) (SMCP-1A).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;

```

RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 176-185.
RP TISSUE-Mast cells;
RC MEDLINE-98343972; PubMed-9677343;
RA McAleese S.M., Pemberton A.D., McGrath M.E., Huntley J.F.,
RA Miller H.R.P.;
RT "Sheep mast-cell proteinases-1 and -3: cDNA cloning, primary
RT structure and molecular modelling of the enzymes and further studies
RT on substrate specificity";
RL Biochem. J. 333:801-809(1998).
RN [2]
RN SEQUENCE OF 20-44.
RA Miller H.R.P., Huntley J.F., Newlands G.F.J.;
RL (in) Caughey G.H. (eds.);
RL Mast cell proteases in immunology and biology, pp.203-235, Marcel
RL Dekker, New York (1995).
RN [3]
RN SEQUENCE OF 20-29.
RC TISSUE-Gastric mucosa;
RX MEDLINE-97184650; PubMed-9032451;
RA Pemberton A.D., Huntley J.F., Miller H.R.P.;
RT "Sheep mast cell proteinase-1: characterization as a member of a new
RT class of dual-specific ruminant chymases.";
RL Biochem. J. 321:665-670(1997).
CC -1- FUNCTION: HAS A CHYMOTRYPSIN-LIKE AND TRYPSIN-LIKE ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- TISSUE SPECIFICITY: MUCOSAL MASP CELLS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. GRANZYME SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: Y14654; CAA74984.1; -.
DR HSSP: P04187; 2CPL.
DR MEROPS: S01.142; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin.1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC.1.
DR PROSITE: PS50240; TRYPSIN_DOM.1.
DR PROSITE: PS00134; TRYPSIN_HIS.1.
DR PROSITE: PS00135; TRYPSIN_SER.1.
KW Hydrolyase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 17
FT PROPEP 18 19
FT CHAIN 20 245
FT ACT_SITE 63 63
FT ACT_SITE 107 107
FT DISULFID 48 64
FT DISULFID 141 207
FT DISULFID 172 186
FT CONFLICT 20 20
FT SEQUENCE 245 AA; 26677 MW; C362CF3367FF8180 CRC64;

Query Match 3.4%; Score 7; DB 1; Length 245;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 40 FULLVAL 46
ID 1111111
AC P76462;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
MCT3_SHEEP
RESULT 7
MCT3_SHEEP

```

```

ID MCT3_SHEEP STANDARD; PRT; 251 AA.
AC 046683;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Mast cell protease 3 precursor (EC 3.4.21.-) (SMCP-3).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OC NCBI_TaxID=9940;
OX 11
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE-Bone marrow;
RC MEDLINE-98343972; PubMed-9677343;
RA McAleese S.M., Pemberton A.D., McGrath M.E., Huntley J.F.,
RA Miller H.R.P.;
RT "Sheep mast-cell proteinases-1 and -3: cDNA cloning, primary
RT structure and molecular modelling of the enzymes and further studies
RT on substrate specificity";
RL Biochem. J. 333:801-809(1998).
CC -1- FUNCTION: HAS A CHYMOTRYPSIN-LIKE AND TRYPSIN-LIKE ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. GRANZYME SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: Y13462; CAA73859.1; -.
DR HSSP: P00763; IDPO.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin.1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC.1.
DR PROSITE: PS50240; TRYPSIN_DOM.1.
DR PROSITE: PS00134; TRYPSIN_HIS.1.
DR PROSITE: PS00135; TRYPSIN_SER.1.
KW Hydrolyase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 17
FT PROPEP 18 19
FT CHAIN 20 251
FT ACT_SITE 63 63
FT ACT_SITE 107 107
FT DISULFID 48 64
FT DISULFID 141 207
FT DISULFID 172 186
FT CARBOHYD 70 70
FT SEQUENCE 251 AA; 27510 MW; DF2F9861144AE909 CRC64;

Query Match 3.4%; Score 7; DB 1; Length 251;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 40 FULLVAL 46
ID 1111111
AC P76462;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
MCT3_SHEEP
RESULT 8
MCT3_SHEEP
ID YFAP_ECOLI STANDARD; PRT; 258 AA.
AC P76462;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
MCT3_SHEEP

```

```

DE Hypothetical protein yfap precursor.
GN yfap OR B2225.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; AE000312; AAC75285.1; -.
DR Ecogene; EG14078; yfap.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 258 HYPOTHETICAL PROTEIN YFAP.
SQ SEQUENCE 258 AA; 28303 MW; 911F2B5C433E5F41 CRC64;
-----
Query Match 3.4%; Score 7; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 41 LLLVALS 47
| | | | |
Db 8 LLLVALS 14
-----
RESULT 9
ID S1C2_HUMAN STANDARD; PRT; 302 AA.
AC 075897;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sulfotransferase 1C2 (EC 2.8.2.-) (SULT1C) (SULT1C#2).
GN SULT1C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=99069375; PubMed=9852044;
RA Sakakibara Y., Yanagisawa K., Katafuchi J., Ringer D.P., Takami Y.,
RA Nakayama T., Suiko M., Liu M.-C.;
RT "Molecular cloning, expression, and characterization of novel human
RT SULT1C sulfotransferases that catalyze the sulfonation of
RT N-hydroxy-2-acetylaminofluorene";
RL J. Biol. Chem. 273:33929-33935(1998).
-----
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20247255; PubMed=10783263;
RA Freimuth R.R., Raftogiannis R.B., Wood T.C., Moon E., Kim U.-J., Xu J.,
RA Siciliano M.J., Weinschlbaum R.M.;
RT "Human sulfotransferases SULT1C1 and SULT1C2: cDNA characterization,
RT gene cloning, and chromosomal localization.";
RL Genomics 65:157-165(2000).
-i- FUNCTION: CATALYZES THE

```

```

CC SULFATE CONJUGATION OF MANY DRUGS, XENOBIOTIC COMPOUNDS, HORMONES,
CC AND NEUROTRANSMITTERS. MAY BE INVOLVED IN THE ACTIVATION OF
CC CARCINOGENIC HYDROXYLAMINES. SHOWS ACTIVITY TOWARDS P-NITROPHENOL
CC AND N-HYDROXY-2-ACETYLAMINO-FLUORENE (N-OH-2AAF).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN FETAL LUNG AND
CC KIDNEY AND AT LOW LEVELS IN FETAL HEART, ADULT KIDNEY, OVARY AND
CC SPINAL CHORD.
CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASES FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; AF055584; AAC95519.1; -.
DR EMBL; AF186263; AAF72810.1; -.
DR HSSP; P50224; 1CJM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransfer; 1.
KW Transferase.
FT CONFLICT 5 5 E -> D (IN REF. 2).
SQ SEQUENCE 302 AA; 35534 MW; DD2F48652E3B56D CRC64;
-----
Query Match 3.4%; Score 7; DB 1; Length 302;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 127 EIQKLAE 133
| | | | |
Db 210 EIQKLAE 216
-----
RESULT 10
ID YRB2_YEAST STANDARD; PRT; 327 AA.
AC P40517;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Ran-specific GTPase-activating protein 2 (Ran binding protein 2)
DE (RANBP2).
GN YRB2 OR YIL063C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RT Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RL [2]
RP CHARACTERIZATION.
RX MEDLINE=98058988; PubMed=9395535;
RA Taura T., Schlenstedt G., Silver P.A.;
RT "Yrb2p is a nuclear protein that interacts with Prp20p, a yeast Rcc1
RT homologue.";
RL J. Biol. Chem. 272:31877-31884(1997).
-----
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98301591; PubMed=9636166;
RA Taura T., Krebber H., Silver P.A.;
RT "A member of the Ran-binding protein family, Yrb2p, is involved in

```

```

RT nuclear protein export."
RL Proc. Natl. Acad. Sci. U.S.A. 95:7427-7432(1998).
CC -1- FUNCTION: Important for the export of protein containing nuclear
CC export signal (NES) out of the nucleus. Stimulates the GTPase
CC activity of GSP1.
CC -1- SUBUNIT: Interacts with GSP1, XPO1 and PRP20.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC -1- SIMILARITY: CONTAINS 1 RANBP1-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z38060; CAA86160.1; -.
DR PIR; S48416; S48416.
DR HSSP; P49792; 1RRP.
DR SGD; S0001325; YRRP.
DR InterPro: IPR000697; RANBP1_WASP.
DR InterPro: IPR000156; RANBP1.
DR Pfam; PF00638; RANBP1; 1.
DR SMART; SM00160; RANBP; 1.
DR GTPase activation; Nuclear protein; Protein transport; Transport.
SQ SEQUENCE 327 AA; 36054 MW; 7A1A626B9171646D CRC64;

Query Match          3.4%; Score 7; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 AKKDTKD 66
    |||||
DB 113 AKKDTKD 119

RESULT 11
GCS1_YEAST          STANDARD;          PRT;          352 AA.
AC P35197;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Zinc finger protein GCS1.
GN GCS1 OR YDL226C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94349929; PubMed-8070409;
RA Ireland L.S., Johnston G.C., Drebot M.A., Dhillon N., Demaggio A.J.,
RA Hoekstra M.F., Singer R.A.;
RT "A member of a novel family of yeast 'zn-finger' proteins mediates
RT the transition from stationary phase to cell proliferation.";
RL EMBO J. 13:3812-3821(1994).

[2]
RP SEQUENCE FROM N.A.
RA Rasmussen S.W.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS A ROLE IN THE RESUMPTION OF YEAST CELL
CC PROLIFERATION FROM STATIONARY PHASE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE GCS1/GLO3/SPS18 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z74274; AAA50389.1; -.
DR EMBL; Z74274; CAA98805.1; -.
DR SGD; S0002385; GCS1.
DR InterPro: IPR001164; Znf-GCS.
DR Pfam; PF01412; ArfGAP; 1.
DR PRINTS; PR00405; REVINTRACTING.
DR SMART; SM00105; ArfGAP; 1.
KW Zinc-finger; Nuclear protein; DNA-binding.
FT ZN_FING 26 49
FT VARIANT 27 27 M->I.
FT VARIANT 29 29 C->Y (IN GCS1-1; SEVERELY AFFECT THE
FT FUNCTION).
SQ SEQUENCE 352 AA; 39296 MW; 3215525F3BF9CF17 CRC64;

Query Match          3.4%; Score 7; DB 1; Length 352;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 TLRGNG 80
    |||||
DB 245 TLRGNG 251

RESULT 12
CPXG_STRSQ          STANDARD;          PRT;          381 AA.
AC P23256;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cytochrome P-450-like protein from a Streptomyces sp.".
GN CYP105C1 OR CHOP.
OS Streptomyces sp.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1931;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90299781; PubMed-2361941;
RA Horii M., Ishizaki T., Paik S.Y., Manome T., Murooka Y.;
RT "An operon containing the genes for cholesterol oxidase and a
RT cytochrome P-450-like protein from a Streptomyces sp.";
RL J. Bacteriol. 172:3644-3653(1990).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31939; AAA26718.1; -.
DR HSSP; O00441; IOXA.
DR InterPro: IPR001128; Cyt_P450.
DR Pfam; PF00067; P450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT BINDING 330 330 HEME (BY SIMILARITY).
SQ SEQUENCE 381 AA; 41703 MW; 841B959C9DDEA59C CRC64;

Query Match          3.4%; Score 7; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 AFLTLVA 45
    |||||
DB 213 AFLTLVA 219

```

```
RESULT 13
YMK7_YEAST
ID YMK7_YEAST STANDARD: PRT: 386 AA.
AC 003760.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 45.2 kDa protein in ZDS2-URAS intergenic region.
GN YML107C OR YM8339.12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; 249210; CA89111.1; -.
DR SGD; S0004575; YML107C.
KW Hypothetical protein.
SO SEQUENCE 386 AA; 45182 MW; 0AC670A62AB2EB56 CRC64;

Query Match 3.4%; Score 7; DB 1; Length 386;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 KEIOKLA 132
    |||||
DB 285 KEIOKLA 291

RESULT 14
NOR_FUSOX
ID NOR_FUSOX STANDARD: PRT: 402 AA.
AC P23295.
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 55A1 (EC 1.14.-.-) (CYPLV1) (P450 DNTR) (Nitric-oxide
DE reductase) (P450 NOR).
GN CYP55A1 OR CYP55.
OS Fusarium oxysporum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Fusarium.
OX NCBI_TaxID=5507;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=MT-811;
RA MEDLINE=91244845; PubMed=2037602;
RA Kitawa H., Tomura D., Oda M., Fukamizu A., Hoshino T., Gotoh O.,
RA Yasui T., Shoun H.;
RT "Nucleotide sequence of the unique nitrate/nitrite-inducible
RT cytochrome P-450 cDNA from Fusarium oxysporum.";
RL J. Biol. Chem. 266:10632-10637(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT-811;
RA MEDLINE=95096031; PubMed=7798191;
RA Tomura D., Obika K., Fukamizu A., Shoun H.;
RT "Nitric oxide reductase cytochrome P-450 gene, CYP 55, of the fungus
RT Fusarium oxysporum containing a potential binding-site for FNR, the
RT transcription factor involved in the regulation of anaerobic growth
```

```
RT of Escherichia coli.";
RL J. Biochem. 116:88-94(1994).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=97475224; PubMed=9334748;
RA Park S.-Y., Shimizu H., Adachi S.-I., Nakagawa A., Tanaka I.,
RA Nakahara K., Shoun H., Obayashi E., Nakamura H., Iizuka T., Shiro Y.;
RT "Crystal structure of nitric oxide reductase from denitrifying fungus
RT Fusarium oxysporum.";
RL Nat. Struct. Biol. 4:827-832(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE=20138220; PubMed=10671516;
RA Shimizu H., Obayashi E., Gomi Y., Arakawa H., Park S.-Y., Nakamura H.,
RA Adachi S.-I., Shoun H., Shiro Y.;
RT "Proton delivery in NO reduction by fungal nitric-oxide reductase.
RT Cryogenic crystallography, spectroscopy, and kinetics of ferric-NO
RT complexes of wild-type and mutant enzymes.";
RL J. Biol. Chem. 275:4816-4826(2000).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX MEDLINE=21159060; PubMed=11258878;
RA Lee D.-S., Park S.-Y., Yamane K., Obayashi E., Hori H., Shiro Y.;
RT "Structural characterization of n-butyl-isocyanide complexes of
RT cytochromes P450nor and P450cam.";
RL Biochemistry 40:2669-2677(2001).
CC -!- FUNCTION: INVOLVED IN A DISSIMILATORY REDUCTION OF NITRITE, ACTS
CC AS A NITRIC OXIDE REDUCTASE.
CC -!- INDUCTION: BY NITRATE/NITRITE.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M63340; AAA33337.1; -.
DR EMBL; D14517; BAA03390.1; -.
DR PIR; A40401; A40401.
DR PIR; JX0335; JX0335.
DR PDB; 1BOM; 15-OCT-97.
DR PDB; 2BOM; 15-OCT-97.
DR PDB; 1CL6; 22-MAR-00.
DR PDB; 1CMJ; 22-MAR-00.
DR PDB; 1CMN; 22-MAR-00.
DR PDB; 1F24; 23-NOV-00.
DR PDB; 1F25; 23-NOV-00.
DR PDB; 1F26; 23-NOV-00.
DR PDB; 1GED; 22-NOV-00.
DR PDB; 1GEI; 29-NOV-00.
DR PDB; 1GEI; 29-NOV-00.
DR InterPro: IPR001128; Cyt_P450.
DR Pfam: PF00067; p450; 1.
DR PRINTS: PR00359; BP450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Heme; Electron transport; 3D-structure.
FT INIT_MET 0
FT BINDING 351 351 HEME.
SO SEQUENCE 402 AA; 44240 MW; 6422AFB24FC7213D CRC64;

Query Match 3.4%; Score 7; DB 1; Length 402;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 AFLVLA 45
    |||||
DB 232 AFLVLA 238

RESULT 15
```

```

NOR1_CYLTO
ID NOR1_CYLTO STANDARD; PRT; 403 AA.
AC 000616;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cytochrome P450 55A2 (EC 1.14.1.1) (Cytochrome P450NOR1).
GN CYP55A2.
OS Cyllindrocarpum tonkinense.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Nectriaceae; Nectria.
OX NCBI_TaxID=42744;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 30561;
RX MEDLINE=97163854; PubMed=9010609;
RA Kudo T., Tomura D., Iiu D.L., Dai X.Q., Shoun H.;
RT "Two isozymes of P450nor of Cyllindrocarpum tonkinense: molecular
RT cloning of the cDNAs and genes, expressions in the yeast, and the
RT putative NAD(P)H-binding site."
RL Biochimie 78:792-799(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; D78511; BAA11408.1; -.
CC DR HSSP; P23295; IF24.
CC DR InterPro; IPR001128; Cyt_P450.
CC DR Pfam; PF00067; P450; 1.
CC DR PRINTS; PR00359; BP450.
CC DR PROSITE; PS00086; CYTOCHROME_P450; FALSE_NEG.
CC DR Oxidoreductase; Monooxygenase; Heme.
CC FT BINDING 352 352 HEME (BY SIMILARITY).
CC SQ SEQUENCE 403 AA; 44273 MW; 1C2AB643FD219534 CRC64;

Query Match 3.48; Score 7; DB 1; Length 403;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 AFLLLVA 45
| | | | |
Db 233 AFLLLVA 239

```

Search completed: October 9, 2002, 16:43:27
 Job time : 15 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:41:17 : Search time 28 seconds

(without alignments)
1272.748 Million cell updates/sec

Title: US-09-674-266a-181

Perfect score: 206

Sequence: 1 RLSCAGTSLSGSGPHPSRRLT.....DFALLDMKKALKLKTLEL 206

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	175	85.0	175	4	O95994 homo sapien
2	34	16.5	175	11	O88312 mus musculu
3	10	4.9	159	13	O90Y05 xenopus lae
4	8	3.9	200	16	O988W1 rhizobium l
5	8	3.9	315	11	O62943 rattus norv
6	8	3.9	637	16	O9RY08
7	8	3.9	1232	5	O960R8
8	8	3.9	1284	5	O9YB20
9	7	3.4	73	12	O91ER9
10	7	3.4	80	8	O47957
11	7	3.4	101	2	O56157
12	7	3.4	101	17	O9Y8V3
13	7	3.4	109	4	O96ZT3
14	7	3.4	112	16	O97SV7
15	7	3.4	114	2	O9KX44
16	7	3.4	114	9	O9XJM3

17	7	3.4	115	16	O91AR5	O914F5 pseudomonas
18	7	3.4	120	3	O05715	O05715 saccharomyc
19	7	3.4	126	17	O97VM7	O97VM7 sulfolobus
20	7	3.4	142	5	O00485	O00485 hydra sp. m
21	7	3.4	167	2	O91435	O91435 salmonella
22	7	3.4	167	5	O9Y6R3	O9Y6R3 drosophila
23	7	3.4	169	17	O96Z48	O96Z48 sulfolobus
24	7	3.4	173	8	O951J1	O951J1 phoxinus er
25	7	3.4	173	8	O951I9	O951I9 phoxinus er
26	7	3.4	185	13	O42251	O42251 xenopus lae
27	7	3.4	188	5	O95XX7	O95XX7 caenorhabdi
28	7	3.4	227	2	O9EMR7	O9EMR7 streptomyce
29	7	3.4	257	10	O9FJ08	O9FJ08 arabidopsis
30	7	3.4	257	12	O9J2L8	O9J2L8 macaca m
31	7	3.4	261	17	O27451	O27451 methanother
32	7	3.4	269	6	O95KW7	O95KW7 bos taurus
33	7	3.4	272	2	O02304	O02304 pseudomonas
34	7	3.4	276	5	P91894	P91894 arenicola m
35	7	3.4	277	10	O9SK11	O9SK11 human
36	7	3.4	291	15	O91UA3	O91UA3 human
37	7	3.4	295	4	O96BG3	O96BG3 homo sapien
38	7	3.4	296	2	O51756	O51756 pseudomonas
39	7	3.4	299	2	O9K212	O9K212 streptomyce
40	7	3.4	299	12	O9WRT4	O9WRT4 macaca m
41	7	3.4	301	16	O9PLM7	O9PLM7 chlamydia m
42	7	3.4	303	11	O9Z248	O9Z248 mus musculu
43	7	3.4	304	4	O96HP9	O96HP9 homo sapien
44	7	3.4	308	5	O9NGT1	O9NGT1 musca domes
45	7	3.4	318	10	O9SJ37	O9SJ37 arabidopsis

ALIGNMENTS

RESULT 1

O95994 PRELIMINARY: PRT: 175 AA.

AC O95994: 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE SECRETED CEMENT GLAND PROTEIN XAG-2 HOMOLOG (ANTERIOR GRADIENT 2

DE (XENOPUS LAEVIS) HOMOLOG).

GN XAG-2/R OR HAG-2/C.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RC TISSUE=BREAST;

RX MEDLINE=9909231; PubMed=9790916;

RA Thompson D.A.; Weigel R.J.;

RT "HAG-2, the human homolog of the Xenopus laevis cement gland gene

RT XAG-2, is coexpressed with estrogen receptor in breast cancer cell

RT lines.";

RL Biochem. Biophys. Res. Commun. 251:111-116(1998).

RN [2]

RN SEQUENCE FROM N.A.

RA Zhang J.S.; Smith D.I.;

RT "Human homolog of XAG is differentially expressed in tumors.";

RL Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.

RN [3]

RN SEQUENCE FROM N.A.

RC TISSUE=COLON ADENOCARCINOMA;

RA Strausberg R.;

RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF038451: AAC82614.1; -

DR EMBL: AF007791: AAC77358.1; -

DR EMBL: AF088867: AAF22484.1; -

DR EMBL: BC015503: AAH15503.1; -

SQ SEQUENCE 175 AA: 19979 MW: F271B1BD377BEE11 CRC64;

Query Match 85.0%; Score 175; DB 4; Length 175;
 Best Local Similarity 100.0%; Pred. No. 2.4e-173;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 32 MEKIPVSAFLLVLTSLTARDTVKPPAKKDTKDSRPKLPTOTSRGMDLIMPTQEE 91
 |||||||
 DB 1 MEKIPVSAFLLVLTSLTARDTVKPPAKKDTKDSRPKLPTOTSRGMDLIMPTQEE 60
 |||||||

Y 92 ALYKSTSNKPLMIHLLDECPHSOALKVFAENKEIOKLAEOFLNLVYETTDKHLSP 151
 |||||||
 DB 61 ALYKSTSNKPLMIHLLDECPHSOALKVFAENKEIOKLAEOFLNLVYETTDKHLSP 120
 |||||||

Y 152 DGQVPRIMEVDPSTLRADITGRYSNRLVAYEPADTALLDNKKALKLKTTEL 206
 |||||||
 DB 121 DGQVPRIMEVDPSTLRADITGRYSNRLVAYEPADTALLDNKKALKLKTTEL 175
 |||||||

RESULT 2
 ID 088312 PRELIMINARY; PRT; 175 AA.
 AC 088312;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE GOB-4 PROTEIN (ANTERIOR GRADIENT 2) (XENOPUS LAEVIS).
 DE LAEVIS).
 GN AGR2 OR GOB-4 OR MAG-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1;TaxID=10090;
 NX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE-INTESTINE;
 RA Komiyama T., Hirohashi S.;
 RT "Cloning of a gene, gob-4, which is expressed in intestinal goblet
 RT cells in mice."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SWISS WEBSTER;
 RA MEDLINE=99009231; PubMed=9790916;
 RA Thompson D.A., Weigel R.J.;
 RT "hcg-2, the human homologue of the Xenopus laevis cement gland gene
 RT XAG-2, is coexpressed with estrogen receptor in breast cancer cell
 RT lines."
 RL Biochem. Biophys. Res. Commun. 251:111-116(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6J; TISSUE-PANCREAS;
 RA MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinaigawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kocha H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schiraldi L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Isono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilmink L.,
 RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RT Nature 409:685-690(2001).
 RN [4]
 RP SEQUENCE FROM N.A.

RC TISSUE-COLON, NORMAL, 5 MONTH OLD MALE MOUSE.;
 RA Strausberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB016592; BAA32044.1; -;
 DR EMBL: AF044262; AAC72705.1; -;
 DR EMBL: AK007677; BAB25181.1; -;
 DR EMBL: BC013354; AAH13354.1; -;
 DR MGD: MGI:1344405; A9r2.
 SQ SEQUENCE 175 AA; 19920 MW; ACC3CFE429B668CA CRC64;

Query Match 16.5%; Score 34; DB 11; Length 175;
 Best Local Similarity 100.0%; Pred. No. 5.1e-27;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 126 KEIOKLAEOFLNLVYETTDKHLSPDGQVPR 159
 |||||||
 DB 95 KEIOKLAEOFLNLVYETTDKHLSPDGQVPR 128
 |||||||

RESULT 3
 ID 090Y05 PRELIMINARY; PRT; 159 AA.
 AC 090Y05;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CEMENT GLAND-SPECIFIC PROTEIN CGS.
 DE Xenopus laevis (African clawed frog).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NC NCB1;TaxID=8355;
 NX [1]
 RP SEQUENCE FROM N.A.
 RA Alexandrova E.M., Novoselov V.V., Zaraisky A.G.;
 RT "Three novel genes expressed in the anterior part of the Xenopus
 RT laevis embryo."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF314056; AAL26844.1; -;
 SQ SEQUENCE 159 AA; 18275 MW; 738D62284838B8BB CRC64;

Query Match 4.9%; Score 10; DB 13; Length 159;
 Best Local Similarity 100.0%; Pred. No. 0.038;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 72 POTSRCGMD 81
 |||||||
 DB 25 POTSRCGMD 34
 |||||||

RESULT 4
 ID 0988W1 PRELIMINARY; PRT; 200 AA.
 AC 0988W1;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
 DE MUR6569 PROTEIN.
 GN MUR6569.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 NC NCB1;TaxID=381;
 NX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF303099;
 RA MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamitsu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

RT Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003009; BAB52836.1; -
 KW Complete proteome.
 SO SEQUENCE 200 AA; 21803 MW; C9CA98E2F8CCF4DE CRC64;

Query Match
 Best Local Similarity 3.9%; Score 8; DB 16; Length 200;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 GDQIMTQ 87
 |||||
 DB 59 GDQIMTQ 66

RESULT 5
 062943 PRELIMINARY; PRT; 315 AA.

AC 062943;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE TASTE_BUD RECEPTOR PROTEIN TB 567.
 GN TB 567.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY;
 RX MEDLINE=97080538; PubMed=8921883;
 RA Thomas M.B., Haines S.L., Akesson R.A.;
 RT "Chemoreceptors expressed in taste, olfactory and male reproductive
 tissues";
 RL Gene 178:1-5(1996).
 DR EMBL: U50948; AAC52910.1; -
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR Pfam: PF00001; 7tm.1; 1.
 DR PROSITE: PS00237; G_PROTEIN_RECIP_FL_1; UNKNOWN_1.
 DR PROSITE: PS50262; G_PROTEIN_RECIP_FL_2; 1.
 KW Receptor.
 SO SEQUENCE 315 AA; 35556 MW; 650D03CD70558652 CRC64;

Query Match
 Best Local Similarity 3.9%; Score 8; DB 11; Length 315;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LLLVALSY 48
 |||||
 DB 142 LLLVALSY 149

RESULT 6
 09RY08 PRELIMINARY; PRT; 637 AA.

AC 09RY08;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE ACYL-COA DEHYDROGENASE, PUTATIVE.
 GN DRA0250.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1199;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-R1;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.O., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,

RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001863; AAF12388.1; -
 DR TIGR: DRA0250; -
 DR InterPro: IPR001552; Acyl-CoA_dh.
 DR Pfam: PF00441; Acyl-CoA_dh; 2.
 DR Pfam: PF02770; Acyl-CoA_dh_M; 1.
 KW Complete proteome.
 SO SEQUENCE 637 AA; 69706 MW; EBCF21B96D1A4E18 CRC64;

Query Match
 Best Local Similarity 3.9%; Score 8; DB 16; Length 637;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 186 ADTALLD 193
 |||||
 DB 422 ADTALLD 429

RESULT 7
 0960R8 PRELIMINARY; PRT; 1232 AA.

AC 0960R8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE LD38671P.
 GN CG5634.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Y, CN BW SP;
 RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Flise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Xu C., Lewis S.E., Rubin G.M., Celinker S.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY051898; AAK93322.1; -
 SO SEQUENCE 1232 AA; 136912 MW; CEF6003D6939A6DD CRC64;

Query Match
 Best Local Similarity 3.9%; Score 8; DB 5; Length 1232;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 RLVAPEPA 186
 |||||
 DB 423 RLVAPEPA 430

RESULT 8
 09VB20 PRELIMINARY; PRT; 1284 AA.

AC 09VB20;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CG5634 PROTEIN.
 GN CG5634.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abell J.F., Abmayyan A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandal D., Bolshakov S.,
 RA Borrova D., Botchan M.R., Bouck J., Brokstam P., Brotler P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodak A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacheb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Syltaks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Maasman D.A., Welnsstock G.M., Weisenbach J.,
 RA Williams S.M., Woodman T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.O.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gdbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL: AEO03761; AAF56723.1; -;
 DR HSSP: P02468; IKLO.
 DR FLYBASE: FBgn0039528; CG5634.
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001798; Kelch.
 DR InterPro: IPR002049; laminin.
 DR InterPro: IPR002165; plexin.
 DR InterPro: IPR003659; PSI.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF01344; Kelch; 6.
 DR Pfam: PF00053; laminin_EGF; 2.
 DR Pfam: PF01437; plexin_repeat; 1.
 DR SMART: SM00042; CUB; 1.
 DR SMART: SM00181; EGF; 1.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00423; PSI; 4.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 1.
 DR EGF-like domain; Glycoprotein.
 SQ SEQUENCE 1284 AA; 142507 MW; B3D559A1D22BF790 CRC64;

Query Match 3.9%; Score 8; DB 5; Length 1284;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 RLVAVEPA 186
 |||||||
 Db 475 RLVAVEPA 482

RESULT 9
 O91ER9
 ID O91ER9 PRELIMINARY; PRT: 73 AA.
 AC O91ER9;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ORF136 SIMILAR TO XCGV ORF13.
 GN ORF136.
 OS Cydia pomonella granulosis virus (CpGV) (Cydia pomonella granulovirus).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
 OX NCBI_TaxID=28289;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MEXICAN 1;
 RX MEDLINE-9318168; PubMed-8445726;
 RA Crook N.E., Clem R.J., Miller L.K.;
 RT "An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif.";
 RL J. Virol. 67:2168-2174(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MEXICAN 1;
 RX MEDLINE-96207404; PubMed-8615018;
 RA Theilmann D.A., Chantler J.K., Stewart S., Flipsen H.T., Viak J.M.,
 RA Crook N.E.;
 RT "Characterization of a highly conserved baculovirus structural protein that is specific for occlusion-derived virions.";
 RL Virology 218:148-158(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MEXICAN 1;
 RX MEDLINE-97380577; PubMed-9237352;
 RA Kang W., Crook N.E., Winstanley D., O'Reilly D.R.;
 RT "Complete sequence and transposon mutagenesis of the BamHI J fragment of Cydia pomonella granulosis virus.";
 RL Virus Genes 14:131-136(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MEXICAN 1;
 RX MEDLINE-98418511; PubMed-9747739;
 RA Kang W., Tristem M., Maeda S., Crook N.E., O'Reilly D.R.;
 RT "Identification and characterization of the Cydia pomonella granulovirus cathepsin and chitinase genes.";
 RL J. Gen. Virol. 79:2283-2292(1998).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MEXICAN 1;
 RA Luque T., Finch R., Crook N., O'Reilly D.R., Winstanley D.;
 RT "The complete sequence of the Cydia pomonella granulovirus genome.";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U53466; AAK70796.1; -;
 SQ SEQUENCE 73 AA; 8302 MW; 7CE181B2ED8B2F5B CRC64;

Query Match 3.4%; Score 7; DB 12; Length 73;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 131 LAEOFVL 137
 |||||||
 Db 27 LAEOFVL 33

RESULT 10
 O47957
 ID O47957 PRELIMINARY; PRT: 80 AA.
 AC O47957;
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE NADH DEHYDROGENASE SUBUNIT 6 (FRAGMENT).

OS Poxinus eos (northern redbelly dace).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Poxinus.
 OX NCBI_TaxID=42663;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96137174; PubMed=85564012;
 RA Tolline C.A.; Baker A.J.;
 RT "Mitochondrial DNA variation and population genetic structure of the
 northern redbelly dace (Poxinus eos).";
 RL Mol. Ecol. 4:745-753(1995).
 DR EMBL; U34407; AAC01504.1; -;
 DR InterPro: IPR001457; oxidored_q3.
 DR Pfam: PF00499; oxidored_q3; 1.
 KW Mitochondrion.
 FT NON_TER 1 1
 FT NON_TER 80 80
 SQ SEQUENCE 80 AA; 8042 MW; D51E54782AD59910 CRC64;

Query Match 3.4%; Score 7; DB 8; Length 80;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 40 FLLVAL 46
 |||||
 DB 5 FLLVAL 11

RESULT 11
 ID 056157 PRELIMINARY; PRT; 101 AA.
 AC 056157;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE C-5 ANTHRONE OXIDASE (FRAGMENT).
 OS Streptomyces venezuelae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=34571;
 RN [1]
 RP SEQUENCE OF 1-93 FROM N.A.
 RC STRAIN-SP5230;
 RX MEDLINE=96146057; PubMed=8581159;
 RA Yang K.; Han L.; Ayer S.W.; Vining L.C.;
 RT "Accumulation of the angucycline antibiotic rabelomycin after
 disruption of an oxygenase gene in the jadomycin B biosynthetic gene
 cluster of Streptomyces venezuelae.";
 RL Microbiology 142:0-0(0).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SP5230;
 RA Yang K.;
 RT Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SP5230;
 RA Yang K.;
 RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SP5230;
 RA Han L.;
 RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN EMBL; U28382; AAC98518.1; -;
 DR InterPro: IPR001064; Crystallin.
 DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 FT NON_TER 101 101
 FT NON_TER 11457 MW; 21DC736567307512 CRC64;
 SQ SEQUENCE 101 AA; 11457 MW; 21DC736567307512 CRC64;

Query Match 3.4%; Score 7; DB 2; Length 101;

Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 200 KILKTEL 206
 |||||
 DB 95 KILKTEL 101

RESULT 12
 ID 09Y8V3 PRELIMINARY; PRT; 101 AA.
 AC 09Y8V3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 10.3 KDA PROTEIN APE2531.
 GN APE2531.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococaceae;
 OC Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y.; Hino Y.; Horikawa H.; Yamazaki S.; Haikawa Y.;
 RA Jin-no K.; Takahashi M.; Sekine M.; Baba S.-I.; Ankal A.; Kosugi H.;
 RA Hosoyama A.; Fukui S.; Nagai Y.; Nishijima K.; Nakazawa H.;
 RA Takamiya M.; Masuda S.; Funahashi T.; Tanaka T.; Kudoh Y.;
 RA Yamazaki J.; Kushida N.; Oguchi A.; Aoki K.-I.; Kubota K.;
 RA Nakamura Y.; Nomura N.; Sako Y.; Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000064; BAA81547.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 101 AA; 10277 MW; BBAF0E9B098E25B3 CRC64;

Query Match 3.4%; Score 7; DB 17; Length 101;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 AGTISGS 11
 |||||
 DB 11 AGTISGS 17

RESULT 13
 ID 09GZT3 PRELIMINARY; PRT; 109 AA.
 AC 09GZT3;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE DC50 (DC23).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Xu X.; Yang Y.; Gao G.; Xiao H.; Chen Z.; Han Z.;
 RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN EMBL; AF271779; AAC44790.1; -;
 DR EMBL; AF253980; AAC44629.1; -;
 DR InterPro: IPR000504; RRM.
 DR Pfam: PF00076; rrm; 1.
 DR SMART: SM00360; RRM; 1.
 DR PROSITE: PSS0102; RRM; 1.
 SQ SEQUENCE 109 AA; 12349 MW; 8BDCB6EA244F17B8 CRC64;

Query Match 3.4%; Score 7; DB 4; Length 109;
 Best Local Similarity 100.0%; Pred. No. 36;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 RPKLPQT 74
 |||||
 Db 95 RPKLPQT 101

RESULT 14

097SV7 PRELIMINARY; PRT; 112 AA.
 AC 097SV7;
 DT 01-OCT-2001 (TREMblrel. 18, Created)
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE CONSERVED DOMAIN PROTEIN.
 GN SP0207.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TIGR4;
 RX MEDLINE-21357209; PubMed-11463916;
 RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Rauden D.,
 RA Holtapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angluoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae";
 RI Science 293:498-506(2001).
 DR EMBL; AE007335; AKK74387.1; -.
 DR TIGR; SP0207; -.
 SQ Complete proteome.
 SQ SEQUENCE 112 AA; 13083 MW; E21BB8790CDE54CE CRC64;

Query Match 3.4%; Score 7; DB 16; Length 112;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 LARDTIV 56
 |||||
 Db 60 LARDTIV 66

RESULT 15

09KXA4 PRELIMINARY; PRT; 114 AA.
 AC 09KXA4;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE HYPOTHETICAL 12.4 KDA PROTEIN.
 GN H0152.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7;
 RX MEDLINE-20198780; PubMed-10734605;
 RA Makino K., Yokoyama K., Kubota Y., Yutsudo C.H., Kimura S.,
 RA Kurokawa K., Ishii K., Hattori M., Tatsuno I., Abe H., Iida T.,
 RA Yamamoto K., Ohnishi M., Hayashi T., Yasunaga T., Honda T.,
 RA Sasaki C., Shinagawa H.;
 RT "Complete nucleotide sequence of the prophage VT2-Sakai carrying the
 verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
 derived from the Sakai outbreak.";

RL Genes Genet. Syst. 74:227-239(1999).
 DR EMBL; AP000422; BAA94180.1; -.
 KW Hypothetical protein.
 SO SEQUENCE 114 AA; 12409 MW; ABF77340011439AB CRC64;

Query Match 3.4%; Score 7; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 KALKLTK 203
 |||||
 Db 108 KALKLTK 114

Search completed: October 9, 2002, 16:44:07
 Job time : 30 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:42:22 : Search time 18 Seconds
(without alignments)
279.537 Million cell updates/sec

Title: US-09-674-266A-181

Perfect score: 206
Sequence: 1 RLSCAGTLCGSGPHPSRLT.....DTALLDMKKALKLTTEL 206

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCtUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	175	85.0	175	4	US-08-916-576B-2
2	175	85.0	175	4	US-09-247-155-106
3	59	28.6	131	4	US-09-247-155-174
4	24	11.7	166	4	US-08-916-576B-6
5	24	11.7	170	4	US-08-916-576B-8
6	7	3.4	183	4	US-08-916-576B-7
7	7	3.4	220	5	US-08-063-552-10
8	7	3.4	220	5	PCR-US93-05704-10
9	7	3.4	411	2	US-08-773-870-1
10	7	3.4	442	3	US-08-993-359-28
11	7	3.4	465	1	US-08-471-496-9
12	7	3.4	465	2	US-08-894-840-9
13	7	3.4	465	3	US-09-139-675-9
14	7	3.4	470	1	US-08-471-496-2
15	7	3.4	470	2	US-08-894-840-2
16	7	3.4	470	3	US-09-139-675-2
17	7	3.4	514	1	US-08-063-552-13
18	7	3.4	514	1	PCR-US93-05704-13
19	7	3.4	515	1	US-08-063-552-4
20	7	3.4	515	5	PCR-US93-05704-4
21	7	3.4	521	1	US-08-063-552-2
22	7	3.4	521	5	PCR-US93-05704-2
23	7	3.4	530	4	US-08-793-044-3
24	6	2.9	21	1	US-08-944-133-13
25	6	2.9	24	4	US-09-326-039-8
26	6	2.9	29	2	US-08-194-981E-39
27	6	2.9	29	2	US-08-194-981E-42

28	6	2.9	29	2	US-08-194-981E-44	Sequence 44, Appl
29	6	2.9	44	1	US-08-530-010-31	Sequence 31, Appl
30	6	2.9	44	2	US-08-484-101B-31	Sequence 31, Appl
31	6	2.9	44	4	US-08-714-524D-31	Sequence 31, Appl
32	6	2.9	55	4	US-09-227-357-216	Sequence 216, App
33	6	2.9	63	4	US-09-326-039-2	Sequence 2, Appl1
34	6	2.9	76	4	US-09-309-487-14	Sequence 14, Appl
35	6	2.9	76	4	US-09-309-487-16	Sequence 16, Appl
36	6	2.9	76	4	US-09-309-487-29	Sequence 29, Appl
37	6	2.9	92	4	US-09-309-487-21	Sequence 21, Appl
38	6	2.9	92	4	US-09-309-487-22	Sequence 22, Appl
39	6	2.9	110	4	US-08-858-207A-273	Sequence 273, App
40	6	2.9	116	3	US-08-545-809A-92	Sequence 92, Appl
41	6	2.9	116	3	US-08-545-809A-118	Sequence 118, App
42	6	2.9	116	3	US-08-545-809A-110	Sequence 110, App
43	6	2.9	117	3	US-08-545-809A-114	Sequence 114, App
44	6	2.9	118	3	US-08-545-809A-116	Sequence 116, App
45	6	2.9	118	3	US-08-545-809A-123	Sequence 123, App

ALIGNMENTS

RESULT 1
US-08-916-576B-2
: Sequence 2, Application US/08916576B
: Patent No. 6171816
: GENERAL INFORMATION:
: APPLICANT: YU, GUO-LIANG
: APPLICANT: DILLON, PATRICK J.
: APPLICANT: EBNERR, REINHARD
: APPLICANT: EBNERR, GREGORY A.
: TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESSES:
: ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
: STREET: 1100 NEW YORK AVENUE, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: US
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/916,576B
: FILING DATE:
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/024,347
: FILING DATE: 23-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: STEFFE, ERIC K.
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 1468.0500001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 175 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-916-576B-2
Query Match 85.0%; Score 175; DB 4; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.3e-159;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 32 MERIPASRLLLVALSYTLARDPTTVKPGAKKDKSRPKLPOTLSRGWGDQILMTQTEEE 91

Db 1 MEKIPVSAFLLLVALSTTLARDTTVKGAKKDKDSRPKLPTQLSRGMDQLMTQYEE 60
QY 92 ALKXSKTSNKPMLTIHHLDECPSQALKKVFKAENKEIOKLAEOFLVNLVYETTDKHLSP 151
Db 61 ALKXSKTSNKPMLTIHHLDECPSQALKKVFKAENKEIOKLAEOFLVNLVYETTDKHLSP 120
QY 152 DGOYVPRIMEVDSLTVRADITGRYSNRLVAYEPADTALLDNMKAKLKLKTEL 206
Db 121 DGOYVPRIMEVDSLTVRADITGRYSNRLVAYEPADTALLDNMKAKLKLKTEL 175

RESULT 2

US-09-247-155-106
Sequence 106, Application US/09247155A
Patent No. 6312922
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, Jean-Baptiste
APPLICANT: Bouquelert, Lydie
TITLE OF INVENTION: Complementary DNAs
FILE REFERENCE: GENSET 021A
CURRENT APPLICATION NUMBER: US/09/247,155A
EARLIER FILING DATE: 1999-02-09
EARLIER APPLICATION NUMBER: 60/074,121
EARLIER FILING DATE: 1998-02-09
EARLIER APPLICATION NUMBER: 60/081,563
EARLIER FILING DATE: 1998-04-13
EARLIER APPLICATION NUMBER: 60/096,116
EARLIER FILING DATE: 1998-08-10
EARLIER APPLICATION NUMBER: 60/099,273
EARLIER FILING DATE: 1998-10-04
NUMBER OF SEQ ID NOS: 182
SOFTWARE: Patent.pm
SEQ ID NO 106
LENGTH: 175
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: -20...-1
US-09-247-155-106

Query Match 85.0%; Score 175; DB 4; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.3e-159;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPVSAFLLLVALSTTLARDTTVKGAKKDKDSRPKLPTQLSRGMDQLMTQYEE 91
Db 1 MEKIPVSAFLLLVALSTTLARDTTVKGAKKDKDSRPKLPTQLSRGMDQLMTQYEE 60
QY 92 ALKXSKTSNKPMLTIHHLDECPSQALKKVFKAENKEIOKLAEOFLVNLVYETTDKHLSP 151
Db 61 ALKXSKTSNKPMLTIHHLDECPSQALKKVFKAENKEIOKLAEOFLVNLVYETTDKHLSP 120
QY 152 DGOYVPRIMEVDSLTVRADITGRYSNRLVAYEPADTALLDNMKAKLKLKTEL 206
Db 121 DGOYVPRIMEVDSLTVRADITGRYSNRLVAYEPADTALLDNMKAKLKLKTEL 175

RESULT 3

US-09-247-155-174
Sequence 174, Application US/09247155A
Patent No. 6312922
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, Jean-Baptiste
APPLICANT: Bouquelert, Lydie
TITLE OF INVENTION: Complementary DNAs
FILE REFERENCE: GENSET 021A
CURRENT APPLICATION NUMBER: US/09/247,155A
CURRENT FILING DATE: 1999-02-09
EARLIER APPLICATION NUMBER: 60/074,121

EARLIER FILING DATE: 1998-02-09
EARLIER APPLICATION NUMBER: 60/081,563
EARLIER FILING DATE: 1998-04-13
EARLIER APPLICATION NUMBER: 60/096,116
EARLIER FILING DATE: 1998-08-10
EARLIER APPLICATION NUMBER: 60/099,273
EARLIER FILING DATE: 1998-10-04
NUMBER OF SEQ ID NOS: 182
SOFTWARE: Patent.pm
SEQ ID NO 174
LENGTH: 131
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: -20...-1
FEATURE:
NAME/KEY: UNSURE
LOCATION: 40,41,43,60,70,76,82,86,105,107
OTHER INFORMATION: Xaa = any one of the twenty amino acids
US-09-247-155-174

Query Match 28.6%; Score 59; DB 4; Length 131;
Best Local Similarity 100.0%; Pred. No. 5.6e-49;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 MEKIPVSAFLLLVALSTTLARDTTVKGAKKDKDSRPKLPTQLSRGMDQLMTQYEE 90
Db 1 MEKIPVSAFLLLVALSTTLARDTTVKGAKKDKDSRPKLPTQLSRGMDQLMTQYEE 59

RESULT 4

US-08-916-576b-6
Sequence 6, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESSES:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 166 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-916-576B-6
Query Match 11.7%; Score 24; DB 4; Length 166;
Best Local Similarity 100.0%; Pred. No. 1.8e-15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 149 LSPDGYVPRIMEVPSLTVRADI 172
|||||
DB 109 LSPDGYVPRIMEVPSLTVRADI 132
RESULT 5
US-08-916-576B-8
Sequence 8, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 170 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-916-576B-8
Query Match 11.7%; Score 24; DB 4; Length 170;
Best Local Similarity 100.0%; Pred. No. 1.9e-15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 149 LSPDGYVPRIMEVPSLTVRADI 172
|||||
DB 112 LSPDGYVPRIMEVPSLTVRADI 135
RESULT 6
US-08-916-576B-7
Sequence 7, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG

APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 183 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-916-576B-7
Query Match 3.4%; Score 7; DB 4; Length 183;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 163 DPSLTVR 169
|||||
DB 140 DPSLTVR 146
RESULT 7
US-08-063-552-10
Sequence 10, Application US/08063552
Patent No. 5688936
GENERAL INFORMATION:
APPLICANT: Edwards, Robert H
TITLE OF INVENTION: Vesicle Membrane Transport Proteins
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheldon & Mak
STREET: 225 South Lake Avenue, Ninth Floor
CITY: Pasadena
STATE: California
COUNTRY: USA
ZIP: 91101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/063,552
FILING DATE: 19930514
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: Farber, Michael B
REGISTRATION NUMBER: 32,612
REFERENCE/DOCKET NUMBER: 9067-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (818) 796-4000
TELEFAX: (818) 795-6321
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Rattus rattus
US-08-063-552-10

Query Match 3.4%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 189 ALLDDNM 195
Db 20 ALLDDNM 26

RESULT 8
PCT-US93-05704-10
Sequence 10, Application PC/TUS9305704
GENERAL INFORMATION:
APPLICANT: Edwards, Robert H
TITLE OF INVENTION: Vesicle Membrane Transport Proteins
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheldon & Mak
STREET: 225 South Lake Avenue, Ninth Floor
CITY: Pasadena
STATE: California
COUNTRY: USA
ZIP: 91001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05704
FILING DATE: 19930611
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Farber, Michael B
REGISTRATION NUMBER: 32,612
REFERENCE/DOCKET NUMBER: 9067-1PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (818) 796-4000
TELEFAX: (818) 795-6321
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Rattus rattus
PCT-US93-05704-10

Query Match 3.4%; Score 7; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 189 ALLDDNM 195
Db 20 ALLDDNM 26

RESULT 9
US-08-773-870-1
Sequence 1, Application US/08773870
Patent No. 5912143
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
TITLE OF INVENTION: NOVEL HUMAN MAGE-LIKE PROTEIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastseq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/773.870
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0179 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 411 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Consensus
CLONE: Consensus
US-08-773-870-1

Query Match 3.4%; Score 7; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 164 PSLTYRA 170
Db 132 PSLTYRA 138

RESULT 10
US-08-993-359-28
Sequence 28, Application US/08993359A
Patent No. 6039942
GENERAL INFORMATION:
APPLICANT: Lassen, Soren F.
APPLICANT: Bech, Lisbeth
APPLICANT: Ohmann, Anders
APPLICANT: Breinholt, Jens C.
APPLICANT: Fugisang, Claus C.
APPLICANT: Ostergaard, Peter R.

;; TITLE OF INVENTION: Phylase Polypeptides
;; FILE REFERENCE: 5383.500-US
;; CURRENT APPLICATION NUMBER: US/08/993,359A
;; CURRENT FILING DATE: 1997-12-18
;; EARLIER APPLICATION NUMBER: 1480/96
;; EARLIER FILING DATE: 1996-12-20
;; EARLIER APPLICATION NUMBER: 1481/96
;; EARLIER FILING DATE: 1996-12-20
;; EARLIER APPLICATION NUMBER: 0301/97
;; EARLIER FILING DATE: 1997-03-18
;; EARLIER APPLICATION NUMBER: 0529/97
;; EARLIER FILING DATE: 1997-05-07
;; EARLIER APPLICATION NUMBER: 1388/97
;; EARLIER FILING DATE: 1997-12-01
;; EARLIER APPLICATION NUMBER: 60/046,082
;; EARLIER FILING DATE: 1997-05-09
;; NUMBER OF SEQ ID NOS: 32
;; SOFTWARE: FASTSEQ for Windows Version 3.0
;; SEQ ID NO: 28
;; LENGTH: 442
;; TYPE: PRT
;; ORGANISM: Paxillus involutus
;; FEATURE:
;; NAME/KEY: SIGNAL
;; LOCATION: (1)...(19)
US-08-993-359-28

Query Match 3.4%; Score 7; DB 3; Length 442;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RLSCT 7
|||||
DB 383 RLSCT 389

RESULT 11
US-08-471-496-9
;; Sequence 9, Application US/08471496
;; Patent No. 5798223
;; GENERAL INFORMATION:
;; APPLICANT: LI, YI
;; APPLICANT: CAO, LIANG
;; APPLICANT: ROSEN, CRAIG
;; TITLE OF INVENTION: HUMAN AMINE TRANSPORTER
;; NUMBER OF SEQUENCES: 9
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;; STREET: 1100 NEW YORK AVENUE, SUITE 600
;; CITY: WASHINGTON
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20005-3934
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/471,496
;; FILING DATE: 06-JUN-1995
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/US95/02645
;; FILING DATE: 01-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: STEFFE, ERIC K
;; REGISTRATION NUMBER: 36,688
;; REFERENCE/DOCKET NUMBER: 1488.0830001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 371-2600
;; TELEFAX: (202) 371-2540
;; INFORMATION FOR SEQ ID NO: 9:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 465 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: not relevant
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-471-496-9

Query Match 3.4%; Score 7; DB 1; Length 465;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 189 ALLDNM 195
|||||
DB 30 ALLDNM 36

RESULT 12
US-08-894-840-9
;; Sequence 9, Application US/08894840
;; Patent No. 5859200
;; GENERAL INFORMATION:
;; APPLICANT: LI, YI
;; APPLICANT: CAO, LIANG
;; APPLICANT: ROSEN, CRAIG A.
;; TITLE OF INVENTION: HUMAN AMINE TRANSPORTER
;; NUMBER OF SEQUENCES: 9
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;; STREET: 1100 NEW YORK AVENUE, SUITE 600
;; CITY: WASHINGTON
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20005-3934
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/894,840
;; FILING DATE: 29-AUG-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: STEFFE, ERIC K
;; REGISTRATION NUMBER: 36,688
;; REFERENCE/DOCKET NUMBER: 1488.0830000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 371-2600
;; TELEFAX: (202) 371-2540
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 465 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: not relevant
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-894-840-9

Query Match 3.4%; Score 7; DB 2; Length 465;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 189 ALLDNM 195
|||||
DB 30 ALLDNM 36

RESULT 13
US-09-139-675-9
;; Sequence 9, Application US/09139675A
;; Patent No. 6117426
;; GENERAL INFORMATION:
;; APPLICANT: LI, YI

APPLICANT: Cao, Liang
APPLICANT: Rosen, Craig
TITLE OF INVENTION: Human Amine Transporter
FILE REFERENCE: 1488.0830003
CURRENT APPLICATION NUMBER: US/09/139,675A
CURRENT FILING DATE: 1998-08-25
EARLIER APPLICATION NUMBER: WO PCT/US95/02645
EARLIER FILING DATE: 1995-03-01
EARLIER APPLICATION NUMBER: US 08/471,496
EARLIER FILING DATE: 1995-06-06
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 9
LENGTH: 465
TYPE: PRT
ORGANISM: Rattus sp.
US-09-139-675-9

Query Match 3.4%; Score 7; DB 3; Length 465;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 ALLDNN 195
DB 30 ALLDNN 36

RESULT 14
US-08-471-496-2
Sequence 2, Application US/08471496
Patent No. 5798223
GENERAL INFORMATION:
APPLICANT: LI, YI
APPLICANT: CAO, LIANG
TITLE OF INVENTION: HUMAN AMINE TRANSPORTER
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,496
FILING DATE: 06-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/02645
FILING DATE: 01-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: STEEFE, ERIC K
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0830001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 470 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-496-2

Query Match 3.4%; Score 7; DB 1; Length 470;
Best Local Similarity 100.0%; Pred. No. 78;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 189 ALLDNN 195
DB 30 ALLDNN 36

RESULT 15
US-08-894-840-2
Sequence 2, Application US/08894840
Patent No. 5859200
GENERAL INFORMATION:
APPLICANT: LI, YI
APPLICANT: CAO, LIANG
APPLICANT: ROSEN, CRAIG A.
TITLE OF INVENTION: HUMAN AMINE TRANSPORTER
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,840
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: STEEFE, ERIC K
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0830000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 470 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-894-840-2

Query Match 3.4%; Score 7; DB 2; Length 470;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 189 ALLDNN 195
DB 30 ALLDNN 36

Search completed: October 9, 2002, 16:45:11
Job time: 19 secs

OM of: US-09-674-266a-181 to: GenEmbl:* out_format: pfs
Date: Oct 8, 2002 10:40 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+ p2n model -DEV=x1h
-O=/cgn2_1/USPRO.spool/US09674266/runat_08102002_093324_26161/app-query.fasta.1.271
-DB=GenEmbl -OPM1=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MAXLEN=LOCAL -OUTTEXT=pfs
-NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09674266.cgn1_1.3973 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPRY -WAIT -THREADS=1
```

Search information block:

Query: US-09-674-266a-181

Database: GenEmbl:*

Database sequences: 1797656

Search length: 187333701

Search time (sec): 1921.540000

Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pat:AX011612	+ 1063.00	1896.06	3.0e-97	1020	AX011612 Sequence 8 from Paten
gb_pat:AX015056	+ 1063.00	1896.06	3.0e-97	1020	AX015056 Sequence 265 from Pat
gb_pr:AF088867	+ 1025.00	1828.39	1.8e-93	956	AF088867 Homo sapiens putative
gb_pr:AF038451	+ 1002.00	1785.87	4.1e-91	1077	AF038451 Homo sapiens secreted
gb_pr:AF115926	+ 1002.00	1781.36	7.4e-91	1701	AF115926 Homo sapiens XAC-2 hc
gb_pr:BC015503	+ 995.00	1775.79	1.5e-80	836	BC015503 Homo sapiens, anterior
gb_pat:AX123828	+ 991.00	1768.15	4.0e-80	875	AX123828 Sequence 1 from patent
gb_pat:AX067336	+ 967.00	1723.45	8.7e-88	778	AX067336 Sequence 40 from Paten
gb_pat:AX341388	+ 964.00	1726.17	1.2e-87	593	AX341388 Sequence 1635 from Pat
gb_pat:AX341084	+ 957.00	1707.74	9.3e-87	1669	AX341084 Sequence 61 from Pat
gb_pat:AX341234	+ 946.00	1690.93	8.0e-86	603	AX341234 Sequence 1481 from Pat
gb_pat:AX177410	+ 914.00	1629.03	2.2e-82	940	AX177410 Sequence 149 from Pat
gb_pat:AX231601	+ 912.00	1630.85	1.8e-82	543	AX231601 Sequence 7 from Patent
gb_pr:AF007791	+ 912.00	1626.57	3.1e-82	838	AF007791 Homo sapiens secreted
gb_pat:AX351468	+ 884.00	1578.03	1.6e-79	699	AX351468 Sequence 215 from Pat
gb_pat:AX341953	+ 876.00	1566.47	6.9e-79	525	AX341953 Sequence 2200 from Pat
gb_pat:AX144213	+ 856.50	1527.36	1.0e-76	793	AX144213 Sequence 78 from patent
gb_pat:AX176400	+ 856.50	1527.36	1.0e-76	793	AX176400 Sequence 78 from patent
gb_r0:AB016592	+ 842.00	1501.74	2.9e-75	758	AB016592 Mus musculus mRNA for
gb_r0:BC044262	+ 842.00	1501.44	2.9e-75	781	BC044262 Mus musculus secreted
gb_r0:BC013334	+ 842.00	1501.18	3.0e-75	802	BC013334 Mus musculus, anterior
gb_pat:AX192616	+ 702.00	1256.17	1.3e-61	409	AX192616 Sequence 183 from Pat
gb_pat:AX341186	+ 691.00	1232.90	2.6e-60	583	AX341186 Sequence 1433 from Pat
gb_pat:AX341116	+ 686.00	1223.83	8.3e-60	588	AX341116 Sequence 1363 from Pat
gb_pat:AX166839	+ 683.00	1222.21	1.0e-59	401	AX166839 Sequence 32 from patent
gb_pat:AX192465	+ 683.00	1222.21	1.0e-59	401	AX192465 Sequence 32 from patent
gb_pat:AX123857	+ 642.00	1148.89	1.2e-55	386	AX123857 Sequence 35 from Patent
gb_pat:AX340984	+ 627.00	1118.97	6.2e-54	554	AX340984 Sequence 1231 from Pat
gb_pat:AX341884	+ 620.00	1105.97	3.1e-53	544	AX341884 Sequence 2131 from Pat
gb_pat:AX341524	+ 616.00	1098.78	7.7e-53	544	AX341524 Sequence 1771 from Pat
gb_pat:AX340668	+ 593.00	1058.03	1.4e-50	512	AX340668 Sequence 915 from Pat
gb_pat:AX341139	+ 589.00	1051.09	3.5e-50	499	AX341139 Sequence 1386 from Pat
gb_pat:AX156194	+ 580.00	1027.34	3.0e-49	529	AX156194 Sequence 51 from Patent
gb_pat:AX341229	+ 576.00	1024.44	7.4e-49	524	AX341229 Sequence 1476 from Pat
gb_pat:AX231609	+ 555.00	987.93	1.2e-46	614	AX231609 Sequence 7 from Patent
gb_pat:AX231718	+ 555.00	987.93	1.2e-46	614	AX231718 Sequence 106 from Pat
gb_pat:AX193063	+ 555.00	987.93	1.2e-46	640	AX193063 Sequence 630 from Pat
gb_pat:AX123830	+ 555.00	986.20	1.4e-46	732	AX123830 Sequence 5 from patent
gb_ov:AF314056	+ 553.00	982.18	2.4e-46	764	AF314056 Xenopus laevis cement

gb_pat:AX302898 - 547.00 979.95 3.2e-46 321 AX302898 Sequence 218 from P
gb_pat:AX322727 + 543.00 972.73 8.1e-46 322 AX322727 Sequence 2 from Pat
gb_pat:AX192553 + 516.00 924.86 3.8e-43 301 AX192553 Sequence 120 from P
gb_pat:AX123858 + 487.50 871.52 3.5e-40 373 AX123858 Sequence 36 from pa
gb_pat:AX341299 + 468.00 835.42 3.6e-38 415 AX341299 Sequence 1546 from

seq_name: gb_pat:AX011612

seq_documentation_block:

LOCUS AX011612 1020 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 8 from Patent W0955858.

ACCESSION AX011612

VERSION AX011612.1 GI:9998136

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 1020)

AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and

Pilarsky,C.

Human nucleic acid sequences obtained from pancreas tumor tissue

Patent WO 955858-A 8 04-NOV-1999;

SCHMITT ARMIN (DE), SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN

BERND (DE); ROSENTHAL ANDRE (DE); METZGEN GRS FUER GENOMFORSCHUN

(DE); PILARSKY CHRISTIAN (DE)

FEATURES

source

1..1020

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 341 a 214 c 218 g 247 t

ORIGIN

alignment_scores:

Quality: 1063.00 Length: 206

Ratio: 5.160 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x AX011612

Align seg 1/1 to: AX011612 from: 1 to: 1020

1 ArgLeuSerGlyAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
|||||
57 AGACTCAGCTGTGCTGCACACTCAGCGCTCTGACCGATCTTACCGC 106
17 GARGLeuThGlnGlyArGTrPValArgLysSerArGValAlaMetGlu 34
|||||
107 CCGACTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 156
34 ysileProValSerAlaPheLeuLeuValAlaLeuSerGlyThrLeu 50
|||||
157 AAATTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 206
51 AlArGAspThrThrValLysProGlyAlaLysLysAspThrLysAsp 67
|||||
207 GCCAGAGATACACACATCTGACCAACCTGAGGCAAAAAGAGACAAAGGACTC 256
67 rArProLysLeuProGlnThrLeuSerArGlyTTrGlyAspGlnLeu 84
|||||
257 TCGACCCAAAGTCCCGACAGCCCTCTCCAGAGGTTGGGGTGACCAACTCA 306
4 leTrpThrGlnThrTyrgLnuAlaLeuTyrlLysSerLysThrSerAsn 100
|||||
307 TCTGACTCAGACATATGAGAACTCTATATATCAAGCAAGCAAGCAAC 356
101 LysProLeuMetIleIleHisHisLeuAspGlnLysProHisSerGlnAl 117
|||||
357 AAACCCCTTGAGATATATTCACACTTGTGATGAGGCCACACAGTCAAGC 406
117 aleuLysLysValPheAlaGlnLysGluIleGlnLysLeuAlaGluG 134

```

|||||
407 TTTAAAGAAAGTGTTCGCTGAAAATAAGAAATCCAGAAATGGCAGAGC 456
134 InPheValLeuAsnLeuValTyrgIuThrThrAspLysHisLeuSer 150
457 AGTTGTCTCTCAATCTGGTTATGAAACACAGACGACAAACCTTTCT 506
151 ProAspGlyGlnTyrgValProArgIleMetPheValAspProSerLeuTh 167
507 CCTATGGCCAGATATGTCGCCAGATTATGTTGTGTGACCATCTCTGAC 556
167 rValArgAlaAspIleThrgIyArGtyrSerAsnArgLeuTyralaTyrg 184
557 AGTTAGAGCCGATATCAGCTGAGAGATATTCAAACCGTCTCTATGCTTACG 606
184 luProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
607 AACCTGCAATACAGCTCTGTTCCTTGACAAACATGAAAGAAAGCTCTCAAG 656
201 LeuLeuLysThrGluLeu 206
657 TTGCTGAAGACTGAATTG 674

```

seq_name: gb_pat:AX015056

seq_documentation_block:

LOCUS AX015056 1020 bp DNA linear PAT 07-SEP-2000

DEFINITION Sequence 265 from Patent WO953040.

ACCESSION AX015056

VERSION AX015056.1 GI:10041195

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1020)

AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and

Pillarsky,C.

TITLE Human nucleic acid sequences from ovarian tumour tissue

JOURNAL Patent: WO 953040-A 265 21-OCT-1999;

SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN

BRAND (DE); ROSENTHAL ANDRE (DE); HINZMANN GES FUER GENOMFORSCHUN

(DE); PILLARSKY CHRISTIAN (DE)

FEATURES

source 1..1020

location/Qualifiers

BASE COUNT 341 a 214 c 218 g 247 t

ORIGIN

Alignment_scores:

Quality: 1063.00 Length: 206

Ratio: 5.160 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:

US-09-674-266a-181 x AX015056 ..

Align seg 1/1 to: AX015056 from: 1 to: 1020

```

1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
|||||
57 AGACTTCAGCTGTGCTGGCAGACTCAGCGGCTCTGGACCGGCATCCATGCGG 106
17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
|||||
107 CCGACTCACACAAAGGACGAGGTGGTGAAGAAATCCAGATGGCATGAGAGA 156
34 yLleProValSerAlaPheLeuLeuValAlaLeuSerTyrrhrLeu 50
|||||
157 AAATTCAGATGTCAGACTTCTGCTCTTGCGCCCTCTCTACACTCTG 206

```

```

51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
|||||
207 GCCAGAGATACACAGCTCAAACTGGAGCCCAAAAAGGACACAAAGGACTC 256
67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
|||||
257 TCGACCCAACTGCCCCAGACCTCTCCAGAGGTGGGGTACCAACAGCA 306
84 lertPrhGlnThrTyrgIuGluAlaLeuTyrgLysSerLysThrSerAsn 100
307 TCTGGACTCAGACATATGAGAGAGCTTATTAATCCAAAGCAAGCAAC 356
101 LysProLeuMetIleIleHisLysLeuAspLysProHisSerGlnAl 117
357 AAACCTTGATGATTAATTCATCCTTGATGATGATGCCACACAGCAAGC 406
117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGlu 134
407 TTTAAAGAAAGTGTTCGCTGAAAATAAGAAATCCAGAAATTTGGCAGAGC 456
134 InPheValLeuAsnLeuValTyrgIuThrThrAspLysHisLeuSer 150
457 AGTTGTCTCTCTCAATCTGGTTATGAAACACAGACAAACACCTTCT 506
151 ProAspGlyGlnTyrgValProArgIleMetPheValAspProSerLeuTh 167
507 CCTGTATGCCAGATATGTCGCCAGATTATGTTGTGTGACCATCTCTGAC 556
167 rValArgAlaAspIleThrgIyArGtyrSerAsnArgLeuTyralaTyrg 184
557 AGTTAGAGCCGATATCAGCTGAGAGATATTCAAACCGTCTCTATGCTTACG 606
184 luProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
607 AACCTGCAATACAGCTCTGTTCCTTGACAAACATGAAAGAAAGCTCTCAAG 656
201 LeuLeuLysThrGluLeu 206
657 TTGCTGAAGACTGAATTG 674

```

seq_name: gb_pr:AF088867

seq_documentation_block:

LOCUS AF088867 956 bp mRNA linear PRI 02-JAN-2000

DEFINITION Homo sapiens putative secreted protein XAG mRNA, complete cds.

ACCESSION AF088867

VERSION AF088867.1 GI:6652811

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 956)

AUTHORS Zhang,J.S. and Smith,D.I.

TITLE Human homolog of XAG is differentially expressed in tumors

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 956)

AUTHORS Zhang,J.S. and Smith,D.I.

TITLE Direct Submission

JOURNAL Submitted (30-AUG-1998) Pathology and Lab. Medicine, Mayo Clinic,

200 SW 1st St., Rochester, MN 55905, USA

Location/Qualifiers

source 1..956

location/Qualifiers

BASE COUNT 148 a 675

ORIGIN

CDS

148..675

SRGMDOLIMTQTYEALYKSKTSNKLPIIHHIDECPSOALKKVFENKEIOKLAE
 QPVLNLVYETTDKHLSPDGOYVPRIMEVDSLVYRADITGRYSNRLAYEPADTALL
 .LDNMKAKLKLTEL"
 BASE COUNT 310 a 202 c 202 g 242 t
 ORIGIN

alignment_scores:
 Quality: 1025.00 Length: 206
 Ratio: 5.025 Gaps: 0
 Percent Similarity: 99.029 Percent Identity: 98.544

alignment_block:
 US-09-674-266a-181 x AF088867 ..
 Align seg 1/1 to: AF088867 from: 1 to: 956

```

1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
|||||
56 AGACTAGCTGCTGCTGCACACTCAGAGCTT.GGACCGCATCTTAGCCG 104
|||||
17 gArGLeuThrGInGlyArGTpValArGlySerArGValAlaMetGluL 34
|||||
105 CCGACTCACAAGAGGAGGTGGGTGAGAAATCCAGAGTTGCCATGGAGA 154
|||||
34 yslLeuProValSerAlaPheLeuLeuValAlaLeuSerTyrrThrLeu 50
|||||
155 AAATTCACATGTCAGACTTCTCTCTGCTGCTGCTCTCTCTCTCTCTG 204
|||||
51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAsp 67
|||||
205 GCCAGAGATACCCACAGTCAAACTGGAGCCAAAAGACCAAGAGACTC 254
|||||
67 rArGProLysLeuProGInThrLeuSerArGlyTrpGlyAspGInLeu 84
|||||
255 TCGACCCAAACTGCCACAGACCTCTCCAGAGGTGGGAGACCAACTCA 304
|||||
84 leTrpThrGInThrTyrgluGluAlaLeuTyrrLysSerLysThrSerAsn 100
|||||
305 TCGGAGCTAGACATATGAAGAGCTCTATATATATCAAGACCAAGCAAC 354
|||||
101 LysProLeuMetIleIleHisLysLeuAspGluCysProHisSerGlnAl 117
|||||
355 AAACCTTGATGATATATCATCTTGATGATGATGCCACACAGTCAAGC 404
|||||
117 aLeuLysLysValPheAlaGluAsnLysGluIleGInLysLeuAlaGlu 134
|||||
405 TTTAAAGAAAGTGTTCGTAATAAATCAAGAAATCCAGAAATGGCAGAGC 454
|||||
134 lPheValLeuLeuAsnLeuValTyrgluThrThrAspLysHisLeuSer 150
|||||
455 AGTTTTCCTCCCTCAATCTGTTTATGAACAACCTGACCAACACCTTCT 504
|||||
151 ProAspGlyGInTyrrValProArGlyIleMetPheValAspProSerLeuTh 167
|||||
505 CCGATGCGCAGATATGCTCCCGAGATTATCTTGTTCACCACTCTCGAC 554
|||||
167 rValAlaGalaAspIleThrGlyArGTyrSerAsnArGLeuTyrrAlaTyrg 184
|||||
555 AGTTAAGCCGATATCATCTGAGAGATATTCAAATCGTCTATAGCTTACG 604
|||||
184 lPheProLysAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
|||||
605 AACCTCAGATTACAGCTCTGTTCTTGACACATGAAGAAAGCTCTCAAG 654
|||||
201 LeuLeuLysThrGluLeu 206
|||||
655 TTGCTGAGAGCTGAATTG 672
|||||

```

seq_name: gb_pr:AF038451
 seq_documentation_block:
 LOCUS AF038451 1077 bp mRNA linear PRI 03-DEC-1998

DEFINITION Homo sapiens, secreted cement gland protein XAG-2 homolog (hAg-2/R)
 mRNA, complete cds.
 ACCESSION AF038451
 VERSION AF038451.1 GI:3779225
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1077)
 AUTHORS Thompson,D.A. and Weigel,R.J.
 TITLE hAg-2, the human homologue of the Xenopus laevis cement gland gene
 XAG-2, is coexpressed with estrogen receptor in breast cancer cell
 lines
 JOURNAL Blochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)
 MEDLINE 99009231
 REFERENCE 2 (bases 1 to 1077)
 AUTHORS Thompson,D.A.
 TITLE Direct Submission
 JOURNAL Submitted (09-DEC-1997) Surgery, Stanford University, 1201 Welch
 Road, MSLS Building, Room P228, Stanford, CA 94305, USA
 FEATURES
 source
 1..1077
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="MCF7"
 /tissue_type="breast"
 /note="Isolated from estrogen receptor-positive breast
 adenocarcinoma cell line"
 1..1077
 /gene="hAg-2/R"
 59..586
 /gene="hAg-2/R"
 /note="Similar to Homo sapiens protein hAg-2/C deposited
 under GenBank Accession Number AF007791, and to Xenopus
 laevis cement gland-specific proteins np77 and XAG-2
 deposited under GenBank Accession Numbers U82110 and
 AF025474, respectively"
 /codon_start=1
 /product="secreted cement gland protein XAG-2 homolog"
 /protein_id="AAC82614.1"
 /db_xref="GI:3779226"
 /translation="MEKIPVSAFULLVALSTLARDTVKRGAKKDTSDRPKLPTL
 SRGMDOLIMTQTYEALYKSKTSNKLPIIHHIDECPSOALKKVFENKEIOKLAE
 QPVLNLVYETTDKHLSPDGOYVPRIMEVDSLVYRADITGRYSNRLAYEPADTALL
 LDNMKAKLKLTEL"

BASE COUNT 367 a 231 c 210 g 269 t
 ORIGIN

alignment_scores:
 Quality: 1002.00 Length: 194
 Ratio: 5.165 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-266a-181 x AF038451 ..
 Align seg 1/1 to: AF038451 from: 1 to: 1077

```

13 ProHisProSerArGArgLeuThrGInGlyArGTpValArGlySerAr 29
|||||
2 CCGGATCTAGCCGCGAGCTCAGACAGGAGGTGGTGAAGAAATCCAG 51
|||||
29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaL 46
|||||
52 AGTTGCCATGAGAGAAATTCACAGTCAATCTTCTCTGTTGGCCC 101
|||||
46 euserTyrrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
|||||
102 TCTCTTACACTCTGGCCAGAGATPACCAAGTCAAACTGGAGCCAAAAG 151
|||||
63 AspThrLysAspSerArGProLysLeuProGInThrLeuSerArGlyTr 79
|||||

```

```

|||||
152 GACACAAAGGACTCTCGACCCAACTGCCAGACCTCTCCAGAGCTTG 201
79  pGlyaspGlnLeuIleTrrPthGlnThrTyrgluGlnuAlaLeuTyrlYsS 96
   |||||||
202 GGGTACCACTCATCTGAGCTGACATATGAAAGACCTCATATTAAT 251
96  eTlysthrSerAsnlySProLeuMetIleIleHisHisLeuaspGluCys 112
   |||||||
252 CCAAGACAAAGCAACACCTTGATGATTAATTCACACTGATGATGTC 301
113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleG1 129
   |||||||
302 CCACACAGTCAGCTTTAAAGAAAGTGTGCTGAATAATAAGAAATCCA 351
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgluThrThra 146
   |||||||
352 GAATTTGGCAGACGAGTTTGTCTCTCATCTGGTTATGCAACACTG 401
146 sPLysHisLeuSerProaspGlyGlnTyValProArgIleMetPheVal 162
   |||||||
402 ACAACACACTTCTCTCGATGGCCAGTATGTCGCCAGATTATGTTGTT 451
163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTySerAsnAr 179
   |||||||
452 GACCCATCTCTGACAGTATGAGCCGATATCACTGAGATATTTCAATCG 501
179 gLeuTyAlaTyrgluProAlaAspThrAlaLeuLeuAsnLeuMet 196
   |||||||
502 TCTCATGCTTACGAACTGCGAGATACAGCTGTGTGCTTGACAACATGA 551
196 ySLysAlaLeuLysLeuLeuLysThrGluLeu 206
   |||||||
552 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 583

seq_name: gb-pr:AF115926

seq_documentation_block:
LOCUS AF115926 1701 bp mRNA linear PRI 30-DEC-2001
DEFINITION Homo sapiens XAG-2 homolog long protein (HPC8) mRNA, complete cds.
ACCESSION AF115926
VERSION AF115926.1 GI:17998664
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1701)
Zhang,J.S. and Smith,D.I.
JOURNAL Unpublished
AUTHORS Identification of human homolog of XAG-2 over-expressed in tumors
2 (bases 1 to 1701)
Zhang,J.S. and Smith,D.I.
JOURNAL Direct Submission
TITLE Submitted (21-DEC-1998) Laboratory of Medicine and Pathology, Mayo
Clinic, 200 SW 1st Street, Rochester, MN 55905, USA
FEATURES
source
1. 1701
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="prostate cancer"
1. 1701
/gene="HPC8"
58. 585
/gene="HPC8"
/gene="HPC8"
/codon_start=1
/product="XAG-2 homolog long protein"
/protein_id="AAL54870.1"
/db_xref="GI:17998665"
/translation="MERIPSAFLLVALSYTLARDPTVYKAGKDKDSDRPKLPQTL
SRGMGOLIMTORFEEALYKSKSNKPMIHHLDRCPSOALKYPAEKETOKLAE
QFYLLLVYTTTKHLSPDGQYPRIMEVDPSSLTVRADITGRISNRLIAYEPADTALL
LDNMKKALKLKTLEL"

```

```

BASE COUNT 557 a 352 c 339 g 453 t
ORIGIN
alignment_scores:
Quality: 1002.00 Length: 194
Ratio: 6.165 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-266a-181 x AF115926
Align seg 1/1 to: AF115926 from: 1 to: 1701
13 ProHisProSerArgArgLeuThrGlnGlyArgTrrValArgLysSerAr 29
|||||
1 CCGCATCCAGCCGCGGACCTCACACAGGACGTGGTGAGAAATCCAG 50
29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaL 46
|||||
51 AGTGCATGAGAGAAATTCAGGTGTACAGATTCTTGCTCTTGAGCCC 100
46 euSerTyThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
|||||
101 TCTCCTACACTCTGGCCAGAGATACACAGTCAAACTGAGCCAAAG 150
63 AspThrLysAspSerArgProLysLeuProGlnThrLysSerArgGlyTr 79
|||||
151 GACACAAAGGACTCTCGACCCAACTGCCAGACCTCTCCAGAGCTTG 200
79 pGlyaspGlnLeuIleTrrPthGlnThrTyrgluGlnuAlaLeuTyrlYsS 96
|||||
201 GGGTACCACTCATCTGAGCTGACATATGAAAGACCTCATATTAAT 250
96 eTlysthrSerAsnlySProLeuMetIleIleHisHisLeuaspGluCys 112
|||||
251 CCAAGACAAAGCAACACCTTGATGATTAATTCACACTGATGATGTC 300
113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleG1 129
|||||
301 CCACACAGTCAGCTTTAAAGAAAGTGTGCTGAATAATAAGAAATCCA 350
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgluThrThra 146
|||||
351 GAATTTGGCAGACGAGTTTGTCTCTCATCTGGTTATGAAACACTG 400
146 sPLysHisLeuSerProaspGlyGlnTyValProArgIleMetPheVal 162
|||||
401 ACAACACACTTCTCTCGATGGCCAGTATGTCGCCAGATTATGTTGTT 450
163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTySerAsnAr 179
|||||
451 GACCCATCTCTGACAGTATGAGCCGATATCACTGGAAGATATTTCAATCG 500
179 gLeuTyAlaTyrgluProAlaAspThrAlaLeuLeuAsnLeuMet 196
|||||
501 TCTCATGCTTACGAACTGCGAGATACAGCTGTGTGCTTGACAACATGA 550
196 ySLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
551 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 582

seq_name: gb-pr:BC015503

seq_documentation_block:
LOCUS BC015503 836 bp mRNA linear PRI 29-OCT-2001
DEFINITION Homo sapiens, anterior gradient 2 (Xenopus laevis) homolog, clone
MGC:9112 IMAGE:3852448, mRNA, complete cds.
ACCESSION BC015503
VERSION BC015503.1 GI:15930124
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens

```

REFERENCE
AUTHORS
TITLE
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 836)

REMARK
COMMENT

Strausberg, R.
Direct Submission
Submitted (01-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mdickpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAK Plate: 20 Row: 0 Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6652811.

FEATURES
source

Location/Qualifiers
1..836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:9112 IMAGE:3852448"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_65"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
57..584
/codon_start=1
/product="anterior gradient 2 (Xenopus laevis) homolog"
/protein_id="AAH1503.1"
/db_xref="GI:15930125"
/translation="MEKIPVSFAFLVALSTYLAADTVKPGAKDKTDSRPKLPQTL SRGMDLIQTFYEALYKSTSNKPLMIITHHDECHSQAALKKVFENKEIOKLAIEQVLLNIYETTDKHLSPDQGVPRIMVDSLVYRADITGYSNRLAYEPAADALL LDNMKAKLKLKTEL"

CDS

BASE COUNT 282 a 181 c 166 g 207 t
ORIGIN

alignment_scores:
Quality: 995.00 Length: 193
Ratio: 5.155 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x BC015503 ..

Align seg 1/1 to: BC015503 from: 1 to: 836

14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
|||||
3 CATCTAGCCGCGGACTCACACAAGCGAGGTGGTGAGAAATCCACAGT 52
|||||
30 lAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuS 47
|||||
53 TGCACATGAGAAAATTCACAGTCTGCTGCTGCTGCTGCTGCT 102
|||||
47 eryThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
|||||
103 CCTACACTCTGGCCAGAGATACACAGTCAAACTGGAGCAAAAAGAC 152
|||||
64 ThrLysSerArgProLysLeuProGlnThrLeuSerArgGlyTrpG1 80

|||||
153 ACAAGACATCTGCAGCCCAACTGCCCCAGACCTCTCCAGAGTTGGG 202

80 yAspGlnLeuIleTrpThrGlnThrTyrgluAlaLeuTyrlsSerL 97
|||||

203 TGACCAACTCATCTGAGCTCAGACATGATGAAGAAGCTTATATAATCA 252

97 ysthrSerAsnLysProLeuMetIleIleHisHisLeuAspGluGysPro 113
|||||

253 AGACAAACCAACAAACCTTGATGATTAATCATCATCTGGATGAGTGCCA 302

114 HisSerGlnAlaLeuLysLysValIlePheAlaGluAsnLysGluLeu 130
|||||

303 CACAGTCACACTTTAAAGAAAGTTGGCTGCAAAATTAACATCCGAA 352

130 sleuAlaGluGlnPheValLeuLeuAsnLeuValTyrgluThrThrAsp 147
|||||

353 ATGGCAGACAGCTTGTCTCCTCATATCTGTTTATGAACAACCTGACA 402

147 ysthisLeuSerProAspGlyGlnTyrrValProArgIleMetPheValAsp 163
|||||

403 AACACCTTCTCTGATGCGCAGATGTCCTCCAGGATTAATGTTGGAC 452

164 ProSerLeuThrValArgAlaAspIleThrGlyArgTrpSerAsnArgLe 180
|||||

453 CCATCTCTGACAGTTAGAGCCGATATCATCTGGAAGATATTCAAACCTCT 502

180 yTyrrAlaTyrgluProAlaAspThrAlaLeuLeuLeuAsnMetLysL 197
|||||

503 CTATGCTTACGAACCTGCAGATACAGCTGTGCTGCTGACACATGANA 552

197 ysaAlaLeuLysLeuLeuLysThrGluLeu 206
|||||

553 AACCTCTCAAGTTGCTGGAAGATCGAATTG 581

seq_name: gb_pat:ARI23828

seq_documentation_block:

LOCUS ARI23828 875 bp DNA linear PAT 16-MAY-2001

DEFINITION Sequence 1 from patent US 6171816.

ACCESSION ARI23828

VERSION ARI23828.1 GI:14109189

KEYWORDS .

SOURCE .

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 875)

AUTHORS Yu, G., Dillon, P. J., Ebner, R. and Endress, G. A.

TITLE Human XAG-1 polynucleotides and polypeptides

JOURNAL Patent: US 6171816-A 1 09-JAN-2001.

FEATURES

source Location/Qualifiers

1..875
/organism="unknown"

BASE COUNT 283 a 189 c 180 g 223 t

ORIGIN

alignment_scores:
Quality: 991.00 Length: 194
Ratio: 5.135 Gaps: 0
Percent Similarity: 99.485 Percent Identity: 99.485

alignment_block:
US-09-674-266A-181 x ARI23828 ..

Align seg 1/1 to: ARI23828 from: 1 to: 875

13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArg 29
|||||
14 CCGATTCCTAGCCGCGCAGATCACACAAGCGAGGTGGTGAGAAATCCAG 63
|||||
29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaL 46
|||||

Percent Similarity: 98.969 Percent Identity: 98.969

alignment_block:
 us-09-674-266a-181 x AX067336 ..

Align seq 1/1 to: AX067336 from: 1 to: 778

```

64 AGTTCAGATGAGAAAAATTCAGTGTGACGATTCCTCTGTCGCC 113
46 euserTythrLeuAlaArgAspThrValLysProGluAlaLys 62
114 TCTCTACACTCTGGCCAGAGATACACAGTCAACCTGGAGCCAAAAG 163
63 AspThrLysAspSerArgProLysLeuProGluThrLeuSerArgLys 79
164 GACACAAAGAGCTCTGACCCAACTGCCAGACCTCTCCAGAGGTG 213
79 pGlyAspGluLeuIleTyrPheGlnThrTyrgLgluAlaLeuTyrLys 96
214 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAACCTCTATATTAAT 263
96 eTyThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluLys 112
264 CCAAGACAGACAAACCCCTTGATGATTTATTCATCCTTGATGAGTGC 313
113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleG1 129
314 CCACACAGTCAAGCTTTAAAGAAAGTGTTCCTGAAATTAAGAAATCCA 363
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgLgluThrPhe 146
364 GAAATTTGCAGAGAGATTGTCTCTCAATCTGTTATGAACAACACTG 413
146 sPLysHisLeuSerProAspGlyGlnTyrgValProArgIleMetPheVal 162
414 ACAAAACACTTTCCTCTGATGCGCCAGTATGTCGCCAGATATATGTTT 463
163 AspProSerLeuThrValAlaArgAlaAspIleThrGlyArgTyrgSerAsn 179
464 GACCATCTCTGACAGTATGAGCCGATATCACTGGAAGATATTCAAATCG 513
179 gLeuTyrrAlaTyrgLupProAlaAspThrAlaLeuLeuLeuAspAsnMet 196
514 TCTCTATGCTTACGAACTCGACAGATACGCTCTGTGCTTGACAAACATGA 563
196 ySLysAlaLeuLysLeuLeuLysThrGluLeu 206
564 AGAAAGCTCTCAAGTGTGTAAGACTGAATTG 595

seq_name: gb_pat:AX067336
seq_documentation_block: 778 bp DNA linear PAT 24-JAN-2001
LOCUS AX067336
DEFINITION Sequence 40 from Patent WO0078960.
ACCESSION AX067336
VERSION AX067336.1 GI:12544960
KEYWORDS
SOURCE
ORGANISM human.
REFERENCE 1 (bases 1 to 778)
AUTHORS Yudin,J. and Milcham,J.L.
TITLE Compositions and methods for the therapy and diagnosis of breast cancer
JOURNAL Patent: WO 0078960-A 40 28-DEC-2000;
CORIXA CORPORATION (US)
FEATURES
Source 1..778
/organism="Homo sapiens"
/db_xref="taxon:9606"
misc_feature 1..778
BASE COUNT 245 a 175 c 158 g 193 t 7 others
ORIGIN

```

Alignment_scores:
 Quality: 967.00 Length: 194
 Ratio: 5.036 Gaps: 0

```

13 ProHisProSerArgArgLeuThrGlnGlyArgTyrValArgLysSer 29
2 CCAGATCCTTACCCCGCCAGCTCACACAAAGGAGGTGGTGAGAAATCCAG 51
29 gValAlaMetClnLysIleProValSerAlaPheLeuLeuValAla 46
52 AGTTGCCATGGAGAAATCCAGTCAAGCATTTCTGCTCTCTTGAGGCC 101
46 euserTythrLeuAlaArgAspThrValLysProGluAlaLys 62
102 TCTCTACACTCTGGCCAGAGATACACAGTCAACCTGGAGCCAAAAG 151
63 AspThrLysAspSerArgProLysLeuProGluThrLeuSerArgLys 79
152 GACACAAAGAGCTCTGACCCAACTGCCAGACCTCTCCAGAGGTG 201
79 pGlyAspGluLeuIleTyrPheGlnThrTyrgLgluAlaLeuTyrLys 96
202 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAACCTCTATATTAAT 251
96 eTyThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluLys 112
252 CCAAGACAGACAAACCCCTTGATGATTTATTCATCCTTGATGAGTGC 301
113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleG1 129
302 CCACACAGTCAAGCTTTAAAGAAAGTGTTCCTGAAATTAAGAAATCCA 351
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgLgluThrPhe 146
352 GAAATTTGCAGAGAGATTGTCTCTCAATCTGTTATGAACAACACTG 401
146 sPLysHisLeuSerProAspGlyGlnTyrgValProArgIleMetPheVal 162
402 ACAAAACACTTTCCTCTGATGCGCCAGTATGTCGCCAGATATATGTTT 451
163 AspProSerLeuThrValAlaArgAlaAspIleThrGlyArgTyrgSerAsn 179
452 GACCATCTCTGACAGTATGAGCCGATATCACTGGAAGATATTCAAATCG 501
179 gLeuTyrrAlaTyrgLupProAlaAspThrAlaLeuLeuLeuAspAsnMet 196
502 TCTCTATGCTTACGAACTCGACAGATCA.GCTCTGTGCTTGACAAACATGA 550
196 ySLysAlaLeuLysLeuLeuLysThrGluLeu 206
551 AGAAAGCTCTCAAGTGTGTAAGACTGAATTG 582

seq_name: gb_pat:AX341388
seq_documentation_block: 593 bp DNA linear PAT 10-JAN-2002
LOCUS AX341388
DEFINITION Sequence 1635 from Patent WO0196388.
ACCESSION AX341388
VERSION AX341388.1 GI:18137370
KEYWORDS
SOURCE
ORGANISM human.
REFERENCE 1 (sites)
AUTHORS Jiang,Y., Harlocker,S.L. and Secrist,H.
TITLE Compositions and methods for the therapy and diagnosis of colon cancer
JOURNAL Patent: WO 0196388-A 1635 20-DEC-2001;
CORIXA CORPORATION (US)

```


FEATURES Location/Qualifiers
source 1..593 /organism="Homo sapiens"
BASE COUNT 188 a 145 c 120 g 138 t 2 others
ORIGIN

alignment_scores:
Quality: 964.00 Length: 189
Ratio: 5.128 Gaps: 0
Percent Similarity: 99.471 Percent Identity: 99.471

alignment_block:
US-09-674-266a-181 x AX341388 ..

Align seg 1/1 to: AX341388 from: 1 to: 593

18 ArgLeuThrGlnGlyArgTyrValArgLysSerArgValAlaMetGluLys 34
|||||
3 CGACTCAGCAAGGCGAGTGGTGAGGAAATCCAGATTGCCATGGAGAA 52
34 s1leProValSerAlaPheLeuLeuValAlaLeuSerTyrThrLeuA 51
|||||
53 AATTCAGTGCAGCATCTCTGCTGTTGGCCCTCTCTACACTCTGG 102
51 laArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSer 67
|||||
103 CCAGAGATCCACAGTCAAACTGGAGCCAAAAGACACAAAGGACTCT 152
68 ArgProLysLeuProGlnThrLeuSerArgGlyTyrGlyAspGlnLeu 84
|||||
153 CCAACCAAACTGCCAGACCTCTCCAGAGGTGGGTGACCAACTCT 202
84 eYrPthrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 101
|||||
203 CTGGACTCAGACATATGAGAAGCTCTATATAATCCAGACAGACAGA 252
101 ysrProLeuMetIleIleHisLysLeuAspGluLysProHisSerGlnAla 117
|||||
253 AACCTTGATGATTAATCATCACTTGGATGAGTCCACACAGTCAAGCT 302
118 LeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGln 134
|||||
303 TTAAAGAAAGGTGTTGCTGAATAATAAGAAATCCAGAAATGGCAGAGA 352
134 nPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 151
|||||
353 GTTGTCCCTCCCAATCTGTTTATGAACAACACTGCACAACTTTCTC 402
151 roAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuThr 167
|||||
403 CTGATGGCCAGTATGTCCTCCAGATTAATGTTGTTGACCATCTGACA 452
168 ValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaArg 184
|||||
453 GTTAGAGCCGATATCACTGGAAGATTAATCAAACTCTATGCTTACGA 502
184 uProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 201
|||||
503 ACCTGCAAGTACAGCTCTGCTTGACACAAACATGAAGAAAGCTCTCAAG 552
201 euLeuLysThrGluLeu 206
|||||
553 TGCCTGAAGACTGAATTG 569

seq_name: gb_pat:ARI77367
seq_documentation_block:
LOCUS ARI77367 1689 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 61 from patent US 6312922.
ACCESSION ARI77367
VERSION ARI77367.1 GI:17919722

KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1689)
AUTHORS Edwards,J.,B.Dumas,Milne., Duclert,A. and Bougueleret,L.
TITLE Complementary DNAs
JOURNAL Patent: US 6312922-A 61 06-NOV-2001;
FEATURES Location/Qualifiers
source 1..1689 /organism="unknown"
BASE COUNT 552 a 350 c 335 g 452 t
ORIGIN

alignment_scores:
Quality: 961.00 Length: 195
Ratio: 5.085 Gaps: 1
Percent Similarity: 96.923 Percent Identity: 96.923

alignment_block:
US-09-674-266a-181 x ARI77367 ..

Align seg 1/1 to: ARI77367 from: 1 to: 1689

12 GlyProHisProSerArgArgLeuThrGlnGlyArgTyrValArgLysSe 28
|||||
9 GGACCCGATCTTAGCCGCGACTCACACAAAGC..... 41
28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValA 45
|||||
42 .AGAGTTCAGATGAGAAATTCAGTGCAGCATCTTGCTCTTGTTGG 90
45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
|||||
91 CCTCTCTTAACACTGGCCAGAGATACCAGTCAAACTGGAGCCAAA 140
62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArg 78
|||||
141 AAGGACCAAAAGGACTCTGCACCCAAAGTCCCAACCTCTCCAGAG 190
78 YrPrgLysAspGlnLeuIleTyrPthrGlnThrTyrGluGluAlaLeuTyr 95
|||||
191 TTGGGGTGACCAACTCATCTGAGCTCAGACATATGAGAAGCTTATTA 240
95 ysrSerLysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGlu 111
|||||
241 AATCCAAGACAAACCAAAACCTTGATGATTAATCATCATCTGGATGAG 290
112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGlu 128
|||||
291 TGCCCAACACAGTCMACTTAAAGAAAGTGTGCTGAATAATAAGAAAT 340
128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThr 145
|||||
341 CCAAGAAATTGGCAGAGAGTGTCTCTCAATCTGGTTATGAACAACA 390
145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
|||||
391 CTGACAAACACCTTTCTCTGATGGCCAGATATGCCAGGATTAATGTT 440
162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSer 178
|||||
441 GTTGACCCCATCTCTGACAGTTAGAGCCGATATCACGGAAGATATCAAA 490
178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsn 195
|||||
491 TCGTCTTAATGCTTACGAACCTGACAGATACAGCTCTGTTGCTTGACAACA 540
195 eLysLysAlaLeuLysLysLeuLysLysThrGluLeu 206
|||||
541 TGAAGAAAGCTCTCAAGTTGCTGAGACAGCTGAATTG 575

```

seq_name: gb_pat:AX341084
seq_documentation_block:
LOCUS AX341084 557 bp DNA linear PAT 10-JAN-2002
DEFINITION Sequence 1331 from Patent WO0196388.
ACCESSION AX341084
VERSION AX341084.1 GI:18137066
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (sites)
AUTHORS Jiang,Y., Harlocker,S.L. and Secrist,H.
TITLE Compositions and methods for the therapy and diagnosis of colon
cancer
JOURNAL Patent: WO 0196388-A 1331 20-DEC-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..557
/organism="Homo sapiens"
/db_xref="taxon:9606"
LOCATION/Qualifiers
BASE COUNT 169 a 143 c 113 g 130 t 2 others
ORIGIN
alignment_scores:
Quality: 957.00 Length: 184
Ratio: 5.201 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-266A-181 x AX341084 ..
Align seg 1/1 to: AX341084 from: 1 to: 557
13 ProHSPoSerArGLeuThrGlnGlyArGTrpValAlaGlySerAr 29
|||||
1 CCGCATCTTACCGCCGACTCAGACAAAGCGAGTGGTGAAGAAATCCAG 50
|||||
29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaL 46
|||||
51 AGTGGCCATGGAGAAATCCAGTGTCCAGCATCTTGTGCTTGTGGCC 100
|||||
46 eUSerTYrThLeuAlaArGspThrThrValLysProGlyAlaLysLys 62
|||||
101 TCTCTTACACTCTGGCCAGAGATACCAAGCAACCTGGAGCCAAAG 150
|||||
63 AsphTrLysAspSerArGProLysLeuProGlnThrLeuSerArGlyTr 79
|||||
151 GACACAAAGGACTCTCGACCCAAACTGCCAGACCTCTCCAGAGTTG 200
|||||
79 pGlyAspGlnLeuIleTrpThrGlnThrTYrGlnGluAlaLeuTYrLys 96
|||||
201 GGGTGACCACTCATCTGGACTCAGACATATGAGAAAGCTCTATTAAT 250
|||||
96 eTYrThSerAsnLysProLeuMetIleIleHSHisLeuAspGluLys 112
|||||
251 CCAAGACAAAGCAACAAACCTTGATGATTATTCATCACTGGATGATGTC 300
|||||
113 ProHISerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleG 129
|||||
301 CCACACACTCAAGCTTTAAAGAAAGTTTGTCTGTAATAAAGAAATCCA 350
|||||
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTYrGlnThrThra 146
|||||
351 GAAATTTGGCAGAGCGATTGTCTCTCAATCTGGTTATGAAACACTG 400
|||||
146 sPLysHisLeuSerProAspGlyGlnTYrValProArgIleMetPheVal 162
|||||
401 ACAAAACACCTTCTCTCATGATGCCAGTATGTCCCAAGGATTTATGTTGT 450
|||||
163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTYrSerAsnAr 179

```

```

|||||
451 GACCATCTCTGACAGCTTAGCGCCGATATCACTGGAGATATTCAAATCG 500
|||||
179 gLeuTYrAlaTYrGlnProAlaAspThrAlaLeuLeuAspAsnMetL 196
|||||
501 TCTTATGCTTAGCAACCTGACAGATACAGCTCTTGTCTGTGACAAATGA 550
|||||
196 ys 196
||
551 AG 552
seq_name: gb_pat:AX341234
seq_documentation_block:
LOCUS AX341234 603 bp DNA linear PAT 10-JAN-2002
DEFINITION Sequence 1481 from Patent WO0196388.
ACCESSION AX341234
VERSION AX341234.1 GI:18137216
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (sites)
AUTHORS Jiang,Y., Harlocker,S.L. and Secrist,H.
TITLE Compositions and methods for the therapy and diagnosis of colon
cancer
JOURNAL Patent: WO 0196388-A 1481 20-DEC-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..603
/organism="Homo sapiens"
/db_xref="taxon:9606"
LOCATION/Qualifiers
BASE COUNT 187 a 153 c 116 g 144 t 3 others
ORIGIN
alignment_scores:
Quality: 946.00 Length: 193
Ratio: 5.059 Gaps: 1
Percent Similarity: 96.891 Percent Identity: 96.373
alignment_block:
US-09-674-266A-181 x AX341234 ..
Align seg 1/1 to: AX341234 from: 1 to: 603
14 HSPoSerArGArGLeuThrGlnGlyArGTrpValAlaGlySerArGVA 30
|||||
3 CATCTTACCGCCGACTCAGACAAAGC.....AGAGT 34
|||||
30 lAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuS 47
|||||
35 TGCCATGGAGAAATTCAGTGTCCAGCATCTTGTGCTTGTGGCCCTCT 84
|||||
47 eTYrThLeuAlaArGspThrThrValLysProGlyAlaLysLysAsp 63
|||||
85 CTTACACTCTGGCCAGAGATACCAAGTCAACCTGGAGCCAAAGAGAC 134
|||||
64 ThrLysAspSerArGProLysLeuProGlnThrLeuSerArGlyTrpG 80
|||||
135 ACAAAAGGACTCTCGACCCAAAGTCCCAAGACCTCTCCAGAGTTGGG 184
|||||
80 yAspGlnLeuIleTrpThrGlnThrTYrGlnGluAlaLeuTYrLysSer 97
|||||
185 TGACCAACTCATCTGGACTCAGACATATGAGAAAGCTCTATATAATCCA 234
|||||
97 ySThSerAsnLysProLeuMetIleIleHSHisLeuAspGluLysPro 113
|||||
235 AGACAAGCAACAAACCTTGATGATTATTCATCACTGGATGATGAGCCCA 284
|||||
114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGln 130
|||||

```

285 CACAGTCAAGCTTTAAAGAAAGTGTGCTGAATAAAGAAATCCAGAA 334
130 sleuAlaGluGlnPheValIleuAsnleuValTyrGluThrAspL 147
335 ATTGGCAGAGCAGTTTGCTCTCATCTGGTTATGAAACACAGTACA 384
147 yshIleuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
385 AACACCTTCTCCTGATGGCCAGATGATGCCCGAGATTTATGTTGTGAC 434
164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
435 CCATCTCTCAGAGTTAGACCCGATATCACTGGAGATATTCAATGCTCT 484
180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
485 GTATGCTTACGAACCTGCAGATACACCTGTTGCTTACCAACATGAAGA 534
197 ysaAlaLeuLysLeuLeuLysThrGluLeu 206
535 AAGCTCTCAAGTTGCTGAAGACTGAATTT 563
seq_name: gb_pat:AR177410
seq_documentation_block:
LOCUS AR177410 940 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 149 from patent US 6312922.
ACCESSION AR177410
VERSION AR177410.1 GI:17919765
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 940)
AUTHORS Edwards,J.,-B.,Dumas.,Mlne., Duclert,A. and Bougueleret,L.
TITLE Complementary DNAs
JOURNAL Patent: US 6312922-A 149 06-NOV-2001;
FEATURES
source location/Qualifiers
1..940
/organism="unknown"
BASE COUNT 304 a 202 c 187 g 233 t 14 others
ORIGIN
alignment_scores:
Quality: 914.00 Length: 207
Ratio: 4.811 Gaps: 2
Percent Similarity: 91.787 Percent Identity: 90.821
alignment_block:
US-09-674-266a-181 x AR177410 ..
Align seg 1/1 to: AR177410 from: 1 to: 940
1 ArgIleuSerCysAlaGlyThrIleuSerGlyProHisProSerArg 17
103 AGACTAGAGTGTGCTGCTGCACACTCAGAGCTT.GGACCGCATCTAGCCG 151
17 gArgIleuThrGlnGlyArgTyrValArgLysSerArgValAlaMetGlnL 34
152 CCGACTCACACAAGGC.....AGAGTTGCCATGGAAA 183
34 ystIleuProValSerAlaPheLeuLeuValAlaLeuSerTyrThrLeu 50
184 AAATTTCAGATGTCAGCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTG 233
51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspS 67
234 GCCAGAGATACCCACAGTCAAACTGAGCCAAAAGACACAAAGAGACTC 283
67 rArgProLysLeuProGlnThrIleuSerArgGlyTyrPglyAspGlnLeu 84
284 TCGACCCAAAGTGCACAGACCTCTCCAGAGGTTGGGGTACCACTCA 333

84 IetPThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
334 TCTGGACACARACATATGAARAACTCTATWTAAATCCARACACCAAC 383
101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
384 AAACCTTGATGATATTCATCCTTGATGADTGGCCACACAGTCAAG 433
117 AleuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGlu 134
434 TTTAAAAAAAKTGTGCTGAAAATTAARAAATCCGAAATTTGGCARANC 483
134 IlnPheValIleuAsnleuValTyrGluThrThrAspLysHisLeuSer 150
484 AGTTGTCCCTCAATCTGTTATGAAACACATGACAAACACTTTCT 533
151 ProAspGlyGlnTyrVal..ProArgIleMetPheValAspProSerLeuT 167
534 CCTGATGGCCAAATATATKTCCTCCCGAGATTTATGTTGTGACCCATCTCTGA 583
167 hrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyr 183
584 CAGTGTAGAGCCGATATCACTGGAAGATATTCAAAYCGTCTCTATGCTTAC 633
184 GluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeu 200
634 GAACCTGCAGATACAGCTCTGTTGCTTGACACATGAAGAAAGCTCTCAA 683
200 sleuLeuLysThrGluLeu 206
684 GTTGCTGAAGACTGAATTTG 702
seq_name: gb_pat:AX231601
seq_documentation_block:
LOCUS AX231601 543 bp DNA linear PAT 11-SEP-2001
DEFINITION Sequence 7 from Patent WO0163290.
ACCESSION AX231601
VERSION AX231601.1 GI:15592500
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 543)
AUTHORS Boyd,R.S., Stamps,A.C., Terrett,J.A. and Tyson,K.L.
TITLE Bcmp-7 as marker for diagnosis of breast cancer
JOURNAL Patent: WO 0163290-A 7 30-AUG-2001;
FEATURES
source location/Qualifiers
1..543
/organism="Homo sapiens"
/db_xref="taxon:9606"
11..538
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC69703.1"
/db_xref="GI:15592501"
/translation="MEKIPVSAFLILVALSYTLARDTVKPKAKKDKRSDSPKLPQTL
SRNGDOLIMQTEALYKSKTSNKLPMITIHNDCHPSQALKKRVAENKIQKLAIE
QFVLNLVETTDKHLSPDGOYVRIMVDSLVYRADITORYSNRLAYRPAQTALL
LDNKKAKLKLKTEL"
BASE COUNT 170 a 135 c 107 g 131 t
ORIGIN
alignment_scores:
Quality: 912.00 Length: 178
Ratio: 5.124 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:

US-09-674-266a-181 x AX231601 ..
Align seg 1/1 to: AX231601 from: 1 to: 543

29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
|||||
2 AGAGTTGGCATTGAGAGAAATCCAGCTGCACATCTTCTGTGTGC 51
45 aLeuSerThrLeuAlaArgAspThrThrValLysProGluAlaLysL 62
|||||
52 CCTCTCTACACTGTGGCCAGAGATACCACAGTCAAACTGGAGCCAAA 101
62 yAspPThrLysAspSerArgProLysLeuProGlnThrLeuSerArgL 78
|||||
102 AGGACACAAAGAGACTCTGCACCAAACTGCCACAGACCTCTCCAGAGT 151
79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrgLysGluAlaLeuTyrl 95
|||||
152 TGGGGTGACCACTCATCTGAGCTCAGACATATGAAGAAGCTATATATA 201
95 sSerLysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluC 112
|||||
202 ATCCAGACAAAGCAACAACCTTGATGATATTCATCATCATTGGATGAGT 251
112 ySPProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
|||||
252 GCCCACACAGTCACAGCTTTAAAGAAAGTGTGCTGAATAATAAGCAATC 301
129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgLutThrTh 145
|||||
302 CAGAAATTGGCAGACAGCTTGTCTCTCTCAATCTGTTTATGAACAAC 351
145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
|||||
352 TGACAAACACCTTCTCTCTGATGGCAGATATGCCAGGATTAATGTTTG 401
162 aLAspProSerLeuThrValAlaArgAlaAspIleThrGlyArgTySerAsn 178
|||||
402 TTGACCCATCTCTGTGACAGTTAGAGCCGATATCATCGAAGAATATTCAA 451
179 ArgLeuTyrlaIeTyrgLutProAlaAspThrAlaLeuLeuAspAsnMe 195
|||||
452 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAAAT 501
195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
502 GAAGAAAGCTCTCAAGTGTGAGAGACTGAAATTG 535

seq_name: gb_pr:AF007791
seq_documentation_block:
LOCUS AF007791 838 bp mRNA linear PRI 14-NOV-1998
DEFINITION Homo sapiens secreted cement gland protein XAG-2 homolog (hAG-2/c)
ACCESSION AF007791
VERSION AF007791.1 GI:3779196
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Thompson,D.A. and Wengcl,R.J.
TITLE hAG-2, the human homologue of the Xenopus laevis cement gland gene
XAG-2, is coexpressed with estrogen receptor in breast cancer cell
lines
JOURNAL Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)
MEDLINE 99009231
REFERENCE 2 (bases 1 to 838)
AUTHORS Thompson,D.A.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1997) Surgery, Stanford, 1201 Welch Road, MSLS
Building, Room P228, Stanford, CA 94305, USA

FEATURES
source location/Qualifiers
1..838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="MCF7; estrogen receptor-expressing"
/cell_type="adenocarcinoma"
/tissue_type="breast"
1..838
/gene="hAG-2/C"
11..538
/gene="hAG-2/C"
/note="similar to Homo sapiens protein hAG-2/R deposited
in GenBank Accession Number AF038451, and to Xenopus
laevis cement gland proteins np77 deposited in GenBank
Accession Number U82110 and XAG-2 deposited in GenBank
Accession Number AF025474"
/codon_start=1
/product="secreted cement gland protein XAG-2 homolog"
/protein_id="AAC7358.1"
/db_xref="GI:3779197"
/translation="MEKIPVSAPFLILVALSYTLARDPTVVKCAKKDKSRPKLPQTL
SRGWSQDILMTQFEALYKSKSNKPLMIIHLDLCPHSQALKKVFAENKEIOKLAE
QVFLNLVETTDKHLSPDQYPRIMFVDPULIVRADITGRSNRLVYEPADTALL
LDNMRKALDKTEL"
BASE COUNT 291 a 170 c 160 g 217 t
ORIGIN

alignment_scores:
Quality: 912.00 Length: 178
Ratio: 5.124 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266a-181 x AF007791 ..
Align seg 1/1 to: AF007791 from: 1 to: 838

29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
|||||
2 AGAGTTGGCATTGAGAGAAATTCAGATGTCAGATTCTTGTCTGTGTGC 51
45 aLeuSerThrLeuAlaArgAspThrThrValLysProGluAlaLysL 62
|||||
52 CCTCTCTACACTGTGGCCAGAGATACCACAGTCAAACTGGAGCCAAA 101
62 yAspPThrLysAspSerArgProLysLeuProGlnThrLeuSerArgL 78
|||||
102 AGGACACAAAGAGACTCTGCACCAAACTGCCACAGTCAAACTCTCCAGAGT 151
79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrgLysGluAlaLeuTyrl 95
|||||
152 TGGGGTGACCACTCATCTGAGCTCAGACATATGAAGAAGCTATATATA 201
95 sSerLysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluC 112
|||||
202 ATCCAGACAAAGCAACAACCTTGATGATATTCATCATCATTGGATGAGT 251
112 ySPProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
|||||
252 GCCCACACAGTCACAGCTTTAAAGAAAGTGTGCTGAATAATAAGCAATC 301
129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgLutThrTh 145
|||||
302 CAGAAATTGGCAGACAGCTTGTCTCTCTCAATCTGTTTATGAACAAC 351
145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
|||||
352 TGACAAACACCTTCTCTGATGGCAGATATGCCAGGATTAATGTTTG 401
162 aLAspProSerLeuThrValAlaArgAlaAspIleThrGlyArgTySerAsn 178
|||||
402 TTGACCCATCTCTGTGACAGTTAGAGCCGATATCATCGAAGAATATTCAAAT 451

```
179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
|||||
452 CGTCTATATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACACAT 501
|||||
195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
502 GAAGAAAGCTCTCAGTTGCTGAAGACTGAATTG 535
|||||
```

OM of: US-09-674-266a-181 to: N_Geneseq_032802:* out_format : pfs
Date: Oct 8, 2002 10:07 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frimet.p2n.model -DEV=xlh
-O/cg2_1/USPTO.spool/US09674266/rnat.08102002_093325_26197/app_query.fasta.1.271
-DB=N_Geneseq_032802 -QEXT=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAEXT=0.050 -XGAPOP=10.000 -XGAEXT=0.500
-FGAPOP=6.000 -FGAEXT=7.000 -YGAPOP=10.000 -YGAEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsuns2
-TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09674266.@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-674-266a-181
Query length: 206
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 213.400000

score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ52865 +		1063.00	2063.38	1.4e-106	10
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ52865 +		1053.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		1043.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		1033.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		1023.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		1013.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		1003.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		993.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		983.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		973.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		963.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		953.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		943.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		933.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		923.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		913.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		903.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		893.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		883.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		873.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		863.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		853.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		843.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		833.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		823.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		813.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		803.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		793.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		783.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		773.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		763.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		753.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		743.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		733.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		723.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		713.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		703.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		693.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		683.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		673.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		663.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		653.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		643.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		633.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		623.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		613.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		603.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		593.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		583.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		573.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		563.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		553.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		543.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		533.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		523.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		513.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		503.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		493.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		483.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		473.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		463.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		453.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		443.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		433.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		423.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		413.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		403.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		393.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		383.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		373.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		363.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		353.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		343.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		333.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		323.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		313.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		303.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		293.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		283.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		273.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		263.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		253.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		243.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		233.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		223.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		213.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		203.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		193.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		183.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		173.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		163.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		153.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		143.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		133.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		123.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		113.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		103.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		93.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		83.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		73.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		63.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		53.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		43.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		33.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		23.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		13.00	1989.96	1.7e-102	92
/SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865 +		3.00	1989.96	1.7e-102	92

/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ52865 - 547.00 1063.77 6.6e-51
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ52865 - 542.00 1042.94 9.6e-50
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAZ52865 - 535.00 1040.71 1.3e-49
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAZ52865 - 529.00 1028.60 6.0e-49
/SID1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ52865 + 516.00 1003.66 1.5e-47

seq_name: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ52865

seq_documentation_block:

ID AAZ52865 standard; CDNA; 1020 BP.

AAZ52865;

14-MAR-2000 (first entry)

Human prostate tumor cDNA library derived EST fragment #8.

Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;

treatment; ds.

Homo sapiens.

DE19820190-A1.

04-NOV-1999.

28-APR-1998; 98DE-1020190.

28-APR-1998; 98DE-1020190.

(META-) METAGEN GES GENOMFORSCHUNG MBH.

Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

WPI; 1999-621386/54.

P-PSDB; AAZ73835, AAZ73836, AAZ73837.

New human nucleic acid sequences from pancreatic tumors, and related

proteins -

Claim 2; Page 188; 502pp; German.

This invention describes novel polypeptides and their encoding nucleic acids derived from human pancreatic tumor tissue which have cytostatic activity. The sequences are also useful in producing pharmaceutical compositions for treatment of pancreatic tumors. AAZ52858-253014 represent expressed sequence tag (EST) fragments derived from a human pancreatic tumor cDNA library and which encode the proteins represented in AAZ73814-Y74252.

Sequence 1020 BP; 341 A; 214 C; 218 G; 247 T; 0 other;

alignment_scores:

Quality: 1063.00 Length: 206

Ratio: 5.160 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

US-09-674-266a-181 x AAZ52865 ..

Align seg 1/1 to: AAZ52865 from: 1 to: 1020

1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
|||||
57 AGACTCAGTGTGCTGACACACACACGCTGACCGATCCTACCG 106
17 gATGLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetCul 34
|||||
107 CCGACTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 156
|||||
34 ystlePovAlSerAlaPheLeuLeuValAlaLeuSerTyrThrLeu 50
|||||

```

157 AAATTCAGTGTACAGATCTTGTGCTGTGTGGCCCTCTCCACACTGTG 206
51 AAlaArgAspThrThrValValLysProGlyAlaLysLysAspThrLysAspSe 67
|||||
207 GCCAGAGTATACACAGTCAAACCTGGAGCCAAAAGGACACAAAGGACTC 256
67 rArgProLysLeuProGlnThrLysSerArgGlyTyrPglLysPglLeuI 84
|||||
257 TCGACCCAAAGTGGCCGACAGCCCTCTCCAGAGTGGGGGTGACCACTCA 306
84 lGtPpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
|||||
307 TGTGACTCAGACATATGAGAAGCTCTATATTAATCCAGACAGCAAC 356
101 LysProLeuMetIleIleHisLysLeuAspGluCysProHisSerGlnAl 117
|||||
357 AAACCTTGAGATATATTCATCATTGAGTGGATGATGCCACAGCAAGC 406
117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGlu 134
|||||
407 TTTAAAGAAAGTGTGTGCTGAATAAAGAAATCCAGAAATTTGGCAGAC 456
134 lPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
|||||
457 AGTTGTCTCTCCTCAATCTGGTTTATGAACAACTGACAAACACTTCT 506
151 ProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuTh 167
|||||
507 CCGATGCGCCAGTATGTCGCCAGCATATATGTTGTTGACCACTCTCTAC 556
167 rValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrG 184
|||||
557 AGTTAGAGCCAGATATCACTGGAGAGATATTCAAACCGCTCTATGCTTAC 606
184 lPheProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
|||||
607 AACCTGCAGATACAGCTCTGTTGCTTGACAACTGAAAGAAAGCTCTCAAG 656
201 LeuLeuLysThrGluLeu 206
|||||
657 TTGCTGAGAGACTGAATTC 674
seq_name: /SIDSI/gcdata/geneseq/geneseqn-emb1/NA2000.DAT: AAC58380
seq_documentation_block:
ID AAC58380 standard; cDNA; 920 BP.
XX
AC AAC58380;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human PRO1030 nucleotide sequence SEQ ID NO:39.
XX
Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
proliferation; tumorigenesis; identification; cancer; cytostatic;
nootropic; neuroprotective; antiinflammatory; immunosuppressive;
immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
neural disorder; glial disorder; astrocytal disorder; angiogenic;
hypothalamic disorder; glandular disorder; macrophagal disorder;
epithelial disorder; stromal disorder; blastocoeleic disorder;
inflammatory disorder; immunologic disorder; ss.
XX
OS Homo sapiens.
XX
MO2000053755-A2.
XX
PD 14-SEP-2000.
XX
PF 06-JAN-2000; 2000MO-US00376.
XX
PR 08-MAR-1999; 99MO-US05028.
PR 02-JUN-1999; 99MO-US12252.
PR 23-JUN-1999; 99US-0141037.

```

```

PR 07-JUL-1999; 99US-0143048.
PR 26-JUL-1999; 99US-0145698.
PR 30-NOV-1999; 99MO-US28313.
PR 20-DEC-1999; 99MO-US30911.
PR 05-JAN-2000; 2000MO-US00219.
XX
PA (GETH ) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA,
PI Watanabe CK, Wood WI;
XX
DR WPI: 2000-572270/53.
XX
PS P-PSDB: AAB24070.
PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT treatment, diagnosis and prevention of cancer -
XX
PS Claim 50; Fig 27; 286bp; English.

```

The present invention describes an isolated antibody that binds to one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535, PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009, PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184, PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094, PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell growth. The PRO polypeptides and nucleotides are useful in the treatment, diagnosis and prevention of cancer. The antibodies and other anti-tumour compounds may be used to treat various conditions, including those characterised by overexpression and/or activation of the amplified CC PRO genes. Exemplary conditions or disorders that are treated with such CC antibodies and other compounds include benign or malignant tumours (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian, colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic carcinomas, sarcomas, glioblastomas, and various head and neck tumours), CC leukemias and lymphoid malignancies, other disorders such as neuronal, CC glial, astrocytal, hypothalamic and other glandular, macrophagal, CC epithelial, stromal and blastocoeleic disorders, and inflammatory, CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR primers and hybridisation probes used in the isolation of the human PRO CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human CC PRO polynucleotide and protein sequences given in the exemplification of the present invention.

Sequence 920 BP; 296 A; 198 C; 195 G; 231 T; 0 other;

alignment_scores: Length: 206
Quality: 1025.00 Gaps: 0
Ratio: 5.025
Percent Similarity: 99.029 Percent Identity: 98.544

alignment_block:
US-09-674-266A-181 x AAC58380 ..

Align seg 1/1 to: AAC58380 from: 1 to: 920

```

1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
|||||
56 AGACTCAGCTGTGCTGCGACACTCAGAAAGCTT.GGACGCGATCTCTAGCGG 104
17 gArgLeuThrGlnGlyArgTyrValArgLysSerArgValAlaMetGluL 34
|||||
105 CCGACTCAGACAGGAGGCGGTGAGAAATCCAGAGTGGCGATGAGAGA 154
34 ySileProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
|||||
155 AAATTCAGTGTACAGATCTTGTGCTGTGTGGCCCTCTCCACACTGTG 204
51 AAlaArgAspThrThrValValLysProGlyAlaLysLysAspThrLysAspSe 67
|||||
205 GCCAGAGATATCCACAGTCAAACCTGGAGCCAAAAGGACACAAAGGACTC 254
67 rArgProLysLeuProGlnThrLysSerArgGlyTyrPglLysPglLeuI 84

```



```

|||||
255 TCGACCAAACTGCCGCCAGACCTCTCCAGAGGTTGGGTCACCAACTCA 304
84 lETpThGlnThrTygluAlaLeuTyLysSerLysThSerAsn 100
305 TCTGACACGACATATCAAGAGCTCTATATTAATCCAGACCAAGCAAC 354
101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnA 117
355 AAACCTTGATGATTAATCATCACTTGATGATGATGATGATGATGATGAT 404
117 AlLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGlu 134
405 TTTAAAGAAAGGTGTTGCTGAAATTAAGAAATCCAGAAATGGCAGAGC 454
134 lnpheValLeuLeuAsnLeuValTyrgLutThrAspLysHisLeuSer 150
455 AGTTGTCTCTCTCAATCTGTTATGAAACAACTGACCAACACCTTCT 504
151 ProAspGlyGlnTyValProArgIleMetPheValAspProSerLeuTh 167
505 CCTGATGCCAGATATGTCCTCCAGGATTTATGTTGTTGACCCATCTCTGAC 554
167 lValArgAlaAspIleThrGlyArgTyLysSerAsnArgLeuTyAlaTyrg 184
555 AGTTAGAGCCGATATCACTGGAAGATATTAATGCTCTATGCTTACG 604
184 lUpProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
605 AACCGCAGATACACGCTCTGTTGCTTGACACATGAGAAAGCTCTCAAG 654
201 LeuLeuLysThrGluLeu 206
655 TTGCTGAAGACTGAATG 672

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAZ77486
seq_documentation_block:
ID AAZ77486 standard; cDNA; 1018 BP.
AC AAZ77486;
XX
XX 10-APR-2000 (first entry)
DE Human ovarian tumor cDNA library derived EST fragment 37.
XX
XX Expressed sequence tag; EST; human; ovarian tumor; anticancer;
XX gene therapy; treatment; ss.
XX
XX Homo sapiens.
XX
XX DE19817557-A1.
XX
XX 21-OCT-1999.
XX
XX 09-APR-1998; 98DE-1017557.
XX
XX 09-APR-1998; 98DE-1017557.
XX
XX 09-APR-1998; 98DE-1017557.
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
XX WPI, 1999-591920/51.
XX
XX P-PSDB; AAY76590.
XX
XX New nucleic acid sequences expressed in ovarian, and some other, cancer
XX tissues, and derived polypeptides, for treatment of ovarian cancer and
XX identification of therapeutic agents -
XX
XX Claim 3; Page 167; 310p; German.
XX
XX This invention describes novel nucleic acid (cDNA) sequences (A) which

```

```

CC have anticancer activity and are highly expressed in ovarian tumor
CC tissue (and some also in testis and breast cancer tissue). The products
CC of the invention can be used for gene therapy. (A) are used (i) for
CC recombinant expression of polypeptides (B) and (ii) to isolate complete
CC genes. (B) are used (i) to identify agents suitable for treatment of
CC ovarian cancer; (ii) directly for treating this form of cancer
CC (including expression from gene therapy vectors) and (iii) for generation
CC of specific antibodies. (A) are identified by assembling ESTs (expressed
CC sequence tags) from a particular tissue type before comparison of
CC expression patterns. This allows a significantly longer fragment of the
CC gene to be revealed, so should reduce the number of failures associated
CC with the fact that ESTs from different libraries may represent different
CC parts of the same unknown gene, distorting the estimated frequency of
CC occurrence in a particular tissue. AAZ77450-277572 represent the human
CC ovarian tumor cDNA library derived EST fragments described in the method
CC of the invention and encode the protein fragments represented in
CC AAY76505-Y76638.
XX
XX
SQ Sequence 1018 BP; 343 A; 212 C; 216 G; 247 T; 0 other;

alignment_scores:
Quality: 1018.50 Length: 207
Ratio: 5.017 Gaps: 2
Percent Similarity: 98.068 Percent Identity: 98.068

alignment_block:
US-09-674-266A-181 x AAZ77486 ..
Align seg 1/1 to: AAZ77486 from: 1 to: 1018

1 ArgLeuSerCysAlaGlyThrLeuSerGlySer.GlyProHisProSerA 17
|||||
57 AGACTGACCTGCTGCGACACTCAGA...AGCTTGACCGCACTCTACGC 103
17 lGArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGlu 33
104 GCGGACTCACACAAAGCAGGTGGTGGAGAAATCCAGAGTTCCCATGGAG 153
34 LysIleProValSerAlaPheLeuLeuValAlaLeuSerTyThrLe 50
154 AAAATTCAGTGTGACGATCTTCTCTGCTGCGCCCTCTCTACACTCT 203
50 uAlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspS 67
|||||
204 GGCAGAGATACACAGTCAAACTGAGCCAAAAGACACAAAGAGACT 253
67 eArgProLysLeuProGlnThrLeuSerArgLysTyProLysAspGlnLeu 83
254 CTGAGACCAAACTGCCCGACCTCTCCAGAGTTGGGCTGACCACTGC 303
84 lETpThGlnThrTygluAlaLeuTyLysSerLysThSerAsn 100
|||||
304 ATCTGACTCAGACATATGAAAGACCTCTATATTAATCCAGCAAGCAA 353
100 nLysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnA 117
354 CAACCTTGATGATTAATCATCACTTGATGATGATGATGATGATGATGAT 403
117 AlLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGlu 133
404 CTTTAAAGAAAGGTGTTGCTGAAATTAAGAAATCCAGAAATGGCAGAG 453
134 GlnPheValLeuLeuAsnLeuValTyrgLutThrAspLysHisLeuSe 150
454 CAGTTGTCTCTCTCAATCTGTTATGAAACAACTGACCAACACCTTTC 503
150 lProAspGlyGlnTyValProArgIleMetPheValAspProSerLeuTh 167
504 TCCTGATGGCCAGTATGTCCTCCAGGATTAATGTTTGTGACCATCTCTGA 553
167 hValArgAlaAspIleThrGlyArgTyLysSerAsnArgLeuTyAlaTy 183
|||||

```

554 CAGTTAGACCGGATATCATCTGGAAGATATTCAAACGCTCTATGCTTAC 603
184 GIUProAlaAspThrAlaLeuLeuLeuAspAsnMetLysAlaLeu 200
604 GAACCTGCACATACGCTGTGCTTGACACATGGAAGAGCTCCAA 653
200 sLeuLeuLysThrGluLeu 206
654 GTTGCTGAAGACTGAATTG 672
seq_name: /STD1/gcgsdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV29047

seq_documentation_block:
ID AAV29047 standard; cDNA; 866 BP.
AC AAV29047;
XX
XX
XX 21-AUG-1998 (first entry)
DE Human protein comprising secretory signal nucleotide sequence 9.
XX
XX Human protein; secretory signal; nutritional source; cytokine;
KW immunity; haematopoiesis; activin; inhibitor; tumour; chemotactic;
KW chemokine; thrombolytic; anti-inflammatory; inhibition;
XX stomach cancer cell; ds.
XX
XX Homo sapiens.
OS
XX
XX
XX Key Location/Qualifiers
XX CDS 73..600
XX FT /*tag= a
XX FT /product= "human protein comprising secretory
XX signal"
XX
XX WO9811217-A2.
XX
XX 19-MAR-1998.
XX
XX 12-SEP-1997; 97WO-JP03239.
XX PF
XX 13-SEP-1996; 96JP-0243060.
XX PR
XX (PROT-) PROTEGENE INC.
XX PA (SAGA) SAGAMI CHEM RES CENTRE.
XX
XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX
XX WPI: 1998-207380/18.
XX DR P-PSDB: AAW37872.
XX
XX Human proteins with secretory signal sequences - used to treat
XX PT immune deficiencies, infections, tumours, and haematopoietic
XX PT disorders, etc.
XX
XX Claim 4; Pages 114-116; 131pp; English.
XX
XX This is the nucleotide sequence of a novel human protein comprising
XX a secretory signal isolated from stomach cancer cells. Its proteins
XX can be used as nutritional sources or supplements. The proteins may
XX also have cytokine functions, immune modulating functions,
XX haematopoiesis regulating activity, activin/inhibin regulating
XX activity, chemotactic/chemokine activity, haemostatic and
XX thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, tumour inhibition activity.
XX
XX Sequence 866 BP; 275 A; 189 C; 178 G; 224 T; 0 other;

alignment_scores:
Quality: 1008.00 Length: 195
Ratio: 5.169 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266a-181 x AAV29047 ..
Align seq 1/1 to: AAV29047 from: 1 to: 866

12 GIUProHisProSerArgArgLeuThrGlnGlyArgThrPValArgLysSe 28
13 GAACCGCATCTAGCCCGCCGACTCACACAAAGGCGAGTGGGTGAGGAATC 62
28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValA 45
63 CAGAGTTGCATGGAGAAAATTCAGGTGACGATTCCTCTCTGTTGG 112
45 IaleuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
113 CCTCTCTTACACTCTGCGCAGAGATACCACTGAACCTGGAGCCAA 162
62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgL 78
163 AAGGACACAAAGAGACTCTCGACCAAACTGCCAGACCTCTCCAGAG 212
78 YTPGlyAspGlnLeuIleThrPheGlnThrTyrGluGlnAlaLeuTyrL 95
213 TTGGGGTGACCAACTCATCTGACACTGACATATGAAGAAGCTATATA 262
95 YSserLysThrSerAsnLysProLeuMetIleIleHisLysAspGlu 111
263 AATCCAAAGACAGCAACAAACCTTGATGATATCTCATCTTGAGAG 312
112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGlu 128
313 TGCCCAACAGACTCAAGCTTAAAGAAAGTGTGCTGAATAATAAGAAAT 362
128 eGlnLysLeuAlaGlnGlnPheValLeuLeuAsnLeuValTyrGluThr 145
363 CCAGAAATGGCAGAGCATTTGCTCTCAANTGTTATGAACAA 412
145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
413 CTGACAAACACCTTCTCTGATGCGCAGATATGCCAGAGATTATGTTT 462
162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSer 178
463 GTTGACCCATCTCTGACAGTTAGAGCCGATATCATCTGGAAGATATCAA 512
178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuAspAsn 195
513 CCGTCTCTATGCTTACGACCTGACAGATACAGCTCTGTGCTTGACACA 562
195 eLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
563 TGAAGAAAGCTCAAGTGTCTGAAAGACTGAATTG 597
seq_name: /STD1/gcgsdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH35026
seq_documentation_block:
ID AAH35026 standard; cDNA; 943 BP.
XX
XX AAH35026;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:2108.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; ss.
XX
XX Homo sapiens.
XX
XX OS
XX
XX PN WO200122920-A2.
XX
XX PD 05-APR-2001.
XX

PF 28-SEP-2000; 2000WO-US26524.
 XX 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR WPI: 2001-235357/24.
 P-PSDB; AAG75621.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PS useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 1; Page 3593; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing of P.
 CC Inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 CC
 SO Sequence 943 BP; 311 A; 199 C; 199 G; 229 T; 5 other;

alignment_scores:
 Quality: 1002.00 Length: 194
 Ratio: 5.165 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-266a-181 x AAH35026 ..

Align seg 1/1 to: AAH35026 from: 1 to: 943

13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValAlaGlySerArg 29
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 26 CCGCATCTCTAGCCGCGACTCACACAAGGCGAGGTGGTGAGAAATCCAG 75
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaLal 46
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 76 AGTTGGCATGGAGAAAATTCACAGTGTCCATCTTCTGCTTGTGCCC 125
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 46 euSerTyrThrLeuAlaArgAspThrThrValLysProGluAlaLysLys 62
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 126 TCCTCTACACTCTGCGCAGATACACAGTCAACCTGGAGCCAAAG 175
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 176 GACACAAAGAGCTCTGACCCCAAGCTGCCCAAGCCCTCTCCAGAGTTG 225
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 79 pgLysArgGlnLeuIleTyrThrGlnThrTyrGluGluAlaLeuTyrLys 96
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 226 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAAGCTTATATAAT 275
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 96 eLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 276 CCAAGACAGCAAAACCCCTTGATGATTAATCATCATCTTGATGATGTC 325

113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleG 129
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 326 CCAACAGCTCAACCTTTAAAGAAAGCTTTGCTGAATAATTAACAATCCA 375
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 146
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 376 GAATATGGCAGACAGATTGTCTCTCCATCTGATGTTTATGAACAAC 425
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 146 sPlyHisLysSerProAspGlyGlnTyrValProArgIleMetPheVal 162
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 426 ACAAAACACCTTCTCTGAGGCCAGATGCTCCCAAGATTAATGTTGTT 475
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 179
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 476 GACCCATCTCTGACAGTTAGAGCCGATATCATCTGAGATATTCAAA 525
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMet 196
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 526 TCCTCATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACACATGA 575
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 196 ySLysAlaLeuLysLeuLeuLysThrGluLeu 206
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 576 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTTG 607
 ||||||||||||||||||||||||||||||||||||||||||||||||||||

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV19155

seq_documentation_block:
 ID AAV19155 standard; cDNA; 875 BP.
 XX
 AC AAV19155;
 XX
 DT 28-AUG-1998 (first entry)
 XX
 DE Human XAG growth factor huXAG-1 cDNA.
 XX
 KW huXAG-1; XAG; growth factor; colon cancer-specific gene;
 KW tumour marker; breast disease; liver disease; lung disease;
 KW emphysema; wound healing; diagnosis; therapy; human; ds.
 XX
 OS Homo sapiens.
 XX
 Key Location/Qualifiers
 FH 71..598
 FT CDS
 FT sig_peptide 71..130
 FT /*tag= b
 FT mat_peptide 131..595
 FT /*tag= c
 FT
 XX
 PN WO9807749-A1.
 XX
 XX 26-FEB-1998.
 PD
 XX
 PF 22-AUG-1997; 97WO-US14139.
 XX
 PR 23-AUG-1996; 96WO-US13766.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Dillon PJ, Edner R, Endress GA, Yu G;
 XX
 DR WPI: 1998-169093/15.
 P-PSDB; AAW37844.
 XX
 XX New isolated human XAG growth factor(s) - used to develop products
 PT for treating e.g. liver, lung or breast diseases or
 PT hyperproliferative disorders, e.g. cancer.
 PS
 PS Claim 5; Fig 1: 141pp; English.
 XX
 CC This cDNA clone codes for huXAG-1 (see AAW37844), a member of a novel
 CC family of human growth factors also including huXAG-2 (see AAW37845)

80 yASpGlnLeuIleTrrpThrGlnThrTyrgluGluAlaLeuTyrlsSerL 97
|
209 TGACCAACTCATCTGAGCTCAGACATATGAGAAGCTCTATATAATCCA 258
97 ySThrSerAsnLysProIeuMetIleIleHisHisLeuAspGluCysPro 113
|
259 AGACACGCAACAAACCTTGATGATTAATTCATCTGATGATGAGGCCCA 308
114 HIsSerGlnAlaLeuLysValPheAlaGluAsnLysGluIleGly 130
|
309 CACACTCAAGCTTTAAAGAAAGTGTCTGAAAATTAAGAAATCCAGAA 358
130 sLeuAlaGluGlnPheValLeuAsnLeuValTyrgluThrThrAspL 147
|
359 ATTGCGACAGAGATTGTCTCTCATCTGTTATGAAACAACTGACA 408
147 ySHIsLeuSerProAspGlyGlnTyrgluValProArgIleMetPheValasp 163
|
409 AACACCTTCTCTGATGGCCAGATATGCCAGGATTATGTGTGAC 458
164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrsAsnArgLe 180
|
459 CCAATCTCGACAGATTAGAGCCGATATCACTGAGATATTCAAATCGTCT 508
180 uTyrgluATyrgluProAlaAspThrAlaLeuLeuAsnLysPasnMetLysL 197
|
509 CTATGCTTACGACACTGCAGATACAGCTGTGTGCTTGACAACATGAMGA 558
197 ySAIaLeuLysLeuLeuLysThrGluLeu 206
|
559 AAGCTCTCAAGTTGCTGAAGACTGAATTG 587
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF44884
seq_documentation_block:
ID AAF44884 standard; cDNA; 778 BP.
XX
AC AAF44884;
XX
XX 28-MAR-2001 (first entry)
DT
XX Human breast cancer related protein coding sequence SEQ ID NO: 40.
DE
XX Human breast cancer; diagnosis; therapy; vaccine; ss.
KW
XX Homo sapiens.
OS
XX WO200078960-A2.
PN
XX 28-DEC-2000.
PD
XX 23-JUN-2000; 2000MO-US17536.
PF
XX 23-JUN-1999; 99DS-0140903.
PR
XX 12-OCT-1999; 99DS-0158980.
PR
XX (CORI-) CORIXA CORP.
PA
XX yuglu J, Mitcham JL;
PI
XX WPI; 2001-041426/05.
DR
XX
PT New polynucleotides encoding breast tumor specific proteins, useful for
PT prevention, treatment and diagnosis of breast cancer -
XX
XX
PS Claim 25; Page 132; 165pp; English.
XX
XX The present invention provides the coding sequences for a number of
CC breast cancer related proteins. These can be used in vaccinations
CC against, diagnosis of and treatment of cancer, particularly breast
CC cancer.
XX

SQ Sequence 778 BP; 245 A; 175 C; 158 G; 193 T; 7 other;
Alignment_scores:
Quality: 967.00 Length: 194
Ratio: 5.036 Gaps: 0
Percent Similarity: 98.969 Percent Identity: 98.969
Alignment_block:
US-09-674-266a-181 x AAF44884 ..
Align seg 1/1 to: AAF44884 from: 1 to: 778
13 ProHisProSerArgArgLeuThrGlnGlyArgTyrgluValArgLysSerar 29
|
2 CCGCATCTTACACCGCCGACATCACACAAAGCAGTGCGAGGAAATCCAG 51
29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaL 46
|
52 AGTGGCCATGAGAAATTCAGGTCCAGATTCCTGCTGCGGCC 101
46 euSerTyThrLeuAlaArgAspThrValLysProGlyValAlasLys 62
|
102 TCTCTTACACTCTGGCCAGASATACACAGTCAAACTGGAGCCAAAAG 151
63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
|
152 GACACAAAGAGACTCGACCCAAACTGCCAGACCTCTCAGAGGTG 201
79 pGlyAspGlnLeuIleTrrpThrGlnThrTyrgluGluAlaLeuTyrlsS 96
|
202 GGGTGACCAACTCATCTGACTCAGACATATGAAGAACTCTATATAAT 251
96 eTyThrSerAsnLysProIeuMetIleIleHisHisLeuAspGluCys 112
|
252 CCAAGACAAAGCAAAACCTTGATGATTAATTCATCACTGATGACGTGC 301
113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleG 129
|
302 CCACACAGTCAAGCTTTAAAGAAAGTGTGCTGAAATAAAGAAATCCA 351
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgluThrThra 146
|
352 GAAATGGCAGAGAGGTTGTCTCTCAATCTGTTATGATAAACACTG 401
146 sPlyHisLeuSerProAspGlyGlnTyrgluValProArgIleMetPheVal 162
|
402 ACAACACACTTCTCTCTGATGCGCCAGTATGTCCAGAGATTATGTTCTT 451
163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrsAsnAr 179
|
452 GACCATCTCTGACAGTATGAGCCGATATCACTGGAATATTTCAAAACG 501
179 gLeuTyrgluATyrgluProAlaAspThrAlaLeuLeuAsnLysMetL 196
|
502 TCTCTATGCTTACGAACTGCAGATCA GCTGTGTGCTTGACACATGA 550
196 ySLySAIaLeuLysLeuLeuLysThrGluLeu 206
|
551 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 582
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ40803
seq_documentation_block:
ID AAZ40803 standard; DNA; 1689 BP.
XX
AC AAZ40803;
XX
XX 18-JAN-2000 (first entry)
DT
XX
XX Secreted protein EST coding sequence 108-008-5-0-A6-FL.
DE
XX
KW Secreted protein; fingerprint identification technique;

KW chromosome mapping: human; hereditary disease; diagnosis; cancer;
 KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
 KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
 KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
 KW hypertension; ss.
 XX
 OS Homo sapiens.
 PN M09940189-A2.
 PD 12-AUG-1999.
 PE 09-FEB-1999; 99WO-IB00282.
 PF 09-FEB-1998; 98US-0074121.
 PR 13-APR-1998; 98US-0081563.
 PR 10-AUG-1998; 98US-0096116.
 PR 04-SEP-1998; 98US-0099273.
 XX
 PA (GEST) GENSET.
 PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
 DR WPI: 1999-600966/51.
 DR P-PSDB; AAY59675.
 XX
 PT Extended CDNA useful for expressing secreted proteins and to obtain
 PT specific antibodies -
 PS
 PS Claim 1; Page 168-169; 244pp; English.
 XX
 CC This sequence encodes a human secreted protein of the invention. The
 CC extended CDNA (or genomic DNA obtainable from them) may be used to
 CC prepare PCR primers and probes. These are useful for forensic matching or
 CC positive identification by DNA sequencing. They may also be used in
 CC alternative fingerprint identification techniques. Antibodies against the
 CC proteins encoded by the extended CDNA are useful in identification of
 CC tissue types or cell species, as well as identifying tissue specific
 CC soluble proteins. The sequences can be used for chromosome mapping and
 CC identification of genes associated with hereditary diseases or drug
 CC response. signal sequences from the CDNA can be used in construction of
 CC secretion vectors. Other sequences derived from the extended CDNA can be
 CC used to clone upstream genomic DNA sequences including promoters. This is
 CC in turn useful for identifying proteins that interact with promoter
 CC sequences. Some of the proteins may be useful in diagnosing and treating
 CC several disorders including, but not limited to: cancer, hyperlipidaemia,
 CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
 CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
 CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.
 XX
 SO Sequence 1689 BP; 552 A; 350 C; 335 G; 452 T; 0 other:
 alignment_scores:
 Quality: 961.00 Length: 195
 Ratio: 5.085 Gaps: 1
 Percent Similarity: 96.923 Percent Identity: 96.923
 alignment_block:
 US-09-674-266A-181 x AA240803 ..
 Align seg 1/1 to: AA240803 from: 1 to: 1689
 12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysse 28
 9 GGACCCGATCTAGCCGCGCCGACTCACACAAGGC..... 41
 28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValA 45
 42 .AGAGTTGCATGGAGAAATTCAGTGTGACATTCCTGCTGTGG 90
 45 ldeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
 |||||||||||||||||||||||||||||||||||||||||
 |||||||||||||||||||||||||||||||||||||||||

91 CCCTCTCTACACTGTGGCCAGATACCAAGTCAACCTGGAGCCAA 140
 62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArg 78
 141 AAGGACACAAGAGCTCGACCCAAACTGCCACAGCCCTCCAGAG 190
 78 yTrpGlyAspGlnLeuIleTrpGlnThrTyrGluGluAlaLeuTyrL 95
 191 TTGGGGTGCACCACTCATCTGACATTCAGATTCGAAAGACTTATTA 240
 95 ySerLeuThrSerAsnLysProLeuMetIleLeuHisLysLeuAspGlu 111
 241 AATCCAGACAGCAACCAACCAACCTTGATGATTTATCATCTGATGAG 290
 112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGlu 128
 291 TGCCCAACACAGTCACAGCTTTAAAGAAAGTGTGGCTGAAAGAT 340
 128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThr 145
 341 CCAGAAATTTGCCAGAGAGCTTGTCTCTCAATCTGGTTTATGAACA 390
 145 hAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
 391 CTGACAAACACACCTTCTCTGATGGCCAGATATGCCAGATATGTT 440
 162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSer 178
 441 GTTGACCCATCTCTGACAGTTAGACCCGATATCAGCAAGATATCA 490
 178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuAsnM 195
 491 TCCTCTGTAGCTTACGAACCTGAGATACAGCTCTGTGCTTGACACA 540
 195 eTrpLysAlaLeuLysLeuLeuLysThrGluLeu 206
 541 TGAAGAAAGCTCTCAAGTTGCTGAGACAGCTGAATTTG 575
 seq_name: /SIBS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA240846
 seq_documentation_block:
 ID AA240846 standard; DNA; 940 BP.
 AC AA240846;
 XX
 DT 18-JAN-2000 (first entry)
 XX
 DE Secreted protein EST coding sequence 78-21-1-B7-FU1.
 XX
 KW Secreted protein; fingerprint identification technique;
 KW chromosome mapping: human; hereditary disease; diagnosis; cancer;
 KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
 KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
 KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
 KW hypertension; ss.
 XX
 OS Homo sapiens.
 PN M09940189-A2.
 PD 12-AUG-1999.
 PE 09-FEB-1999; 99WO-IB00282.
 PF 09-FEB-1998; 98US-0074121.
 PR 13-APR-1998; 98US-0081563.
 PR 10-AUG-1998; 98US-0096116.
 PR 04-SEP-1998; 98US-0099273.
 XX
 PA (GEST) GENSET.
 PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
 XX

DR WPI; 1999-600966/51.
DR P-PSDB; AAY59718.

XX Extended cDNAs useful for expressing secreted proteins and to obtain
PT specific antibodies -
PS Claim 1: Page 226; 244pp; English.

XX This sequence encodes a human secreted protein of the invention. The
CC extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC prepare PCR primers and probes. These are useful for forensic matching or
CC positive identification by DNA sequencing. They may also be used in
CC alternative fingerprint identification techniques. Antibodies against the
CC proteins encoded by the extended cDNAs are useful in identification of
CC tissue types or cell species, as well as identifying tissue specific
CC soluble proteins. The sequences can be used for chromosome mapping and
CC identification of genes associated with hereditary diseases or drug
CC response. Signal sequences from the cDNAs can be used in construction of
CC secretion vectors. Other sequences derived from the extended cDNAs can be
CC used to clone upstream genomic DNA sequences including promoters. This is
CC in turn useful for identifying proteins that interact with promoter
CC sequences. Some of the proteins may be useful in diagnosing and treating
CC several disorders including, but not limited to: cancer, hyperlipidaemia,
CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.

XX Sequence 940 BP; 304 A; 202 C; 187 G; 233 T; 14 other;

alignment_scores:
Quality: 914.00 Length: 207
Ratio: 4.811 Gaps: 2
Percent Similarity: 91.787 Percent Identity: 90.821

alignment_block:
US-09-674-266A-181 x AA240846 ..

Align seg 1/1 to: AA240846 from: 1 to: 940

1 ArgLeuSerGlyAlaGlyThrLeuSerGlyProHisProSerAr 17
103 AGACTGACCTGCTGCGCACACTCGAAGCTT.GGACCGCATCTGAGCCG 151
17 GargLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
152 CCGACTCACACAAGC.....AGACTGCGCATGAGAA 183
34 ysIleProValSerAlaPheLeuLeuValAlaLeuSerTrpThrLeu 50
184 AAATTCCAGTGCAGCATCTGCTCTGTGGCCCTCTCTACACTCTG 233
51 AlaArgAspTrpThrValLysProGlyAlaLysLysAspTrpLysAspSe 67
234 GCGAGAGATCACACAGTCAAACTGGAGCCAAAAGACACAAAGGACTC 283
67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeu 84
284 TCGACCCAAACTGCCCGACACCTCTCCAGAGGTGGGGTGACCAACTCA 333
84 leTrpThrGlnThrTyrgLugLualalaLeuTyrlYrLysSerLysThrSerAsn 100
334 TCTGGACACARACATATGAAARAARCTCTATWTAAATCCAAARCAAGCAAC 383
101 LysProLeuMetIleIleHisHisLeuAspGlyLysProHisSerGlnAl 117
384 AAACCTTGATGATATTCATCATTGATGATGATGATGATGATGATGATGAT 433
117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGlu 134
434 TTTTAAAAAAATGTTGCTGAATAAARAATCCAGAAATTCGACARANC 483
134 InPheValLeuLeuAsnLeuValTyrgLutThrAspLysHisLeuSer 150

|||||
484 AGTTGTCCCTCCATCTGGTTATGAAACACGACAAACACCTTCT 533
151 ProAspGlyGlnTrpVal.ProArgIleMetPheValAspProSerLeuT 167
534 CCGTGGCCAAATATKTCCTCCCGGATTTGTTGTTGATCCCATCTCGA 583
167 hrValArgAlaAspIleThrGlyArgTySerAsnArgLeuTyraLayr 183
584 CAGTTAGAGCCGATATCACTGGAGATATTCAAAVCGTCTGTAGGCTTAC 633
184 GluProAlaAspTrpThrAlaLeuLeuAspAsnMetLysLysAlaLeuLy 200
634 GAACTCGCAGATACAGCTCTGTGCTTGACAAATGAGAAAGCTCTCA 683
200 sLeuLeuLysThrGluLeu 206
684 GTTGCTGAAGACTGAAATTG 702

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAS13480

seq_documentation_block:

ID AAS13480 standard; DNA: 543 BP.

AC AAS13480;

DT 19-DEC-2001 (first entry)

DE DNA encoding breast cancer cell membrane protein 7 (BCMP 7).

XX Breast cancer cell membrane protein 7; BCMP 7; breast cancer;

KW metastasis; cytosolic.; gene therapy; antibody; antisense;

KW vaccine; chromosome 7p21.3; ds.

XX Homo sapiens.

OS Key Location/Qualifiers

FT CDS 11..538
FT /*lag- a
FT /product- "BCMP 7"

FT sig_peptide 11..70
FT /*lag- b

FT mat_peptide 71..535
FT /*lag- c
FT /note- "Mature BCMP 7"

XX MO200163290-A1.

XX 30-AUG-2001.

XX 21-FEB-2001; 2001MO-GB00734.

XX 25-FEB-2000; 2000GB-0004576.

XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX Boyd RS, Stamps AC, Terrett JA, Tyson KL;

XX WPI: 2001-570651/64.

XX P-PSDB: AAU08804.

PT Diagnosing, preventing and treating breast cancer using a breast cancer

PS cell membrane protein BCMP 7 -

PS Claim 6; Fig 1; 62pp; English.

CC The invention describes the novel use of a protein found in breast
CC cancer cell membranes (BCMP 7) for diagnosing, preventing and treating
CC breast cancers. The peptide has cytosolic action and potential uses in
CC gene therapy and in vaccines. The polypeptide, antisense nucleic acids,
CC or fusion proteins comprising and Green Fluorescent protein or the DsRed
CC fluorescent protein, antibodies specific for and/or nucleic acid are

CC used for the prevention and/or treatment of breast cancer. Antibodies
 CC against may also be used for screening for and/or diagnosis of breast
 CC cancer in a patient. The method for monitoring/assessing breast cancer
 CC treatment in a patient and for the identification of metastatic breast
 CC cancer cells in samples from a patient. This sequence encodes breast
 CC cancer cell associated protein 7 (BCMP 7), located on chromosome
 CC 7p21.3, described in the method of the invention.

XX
 SQ Sequence 543 BP; 170 A; 135 C; 107 G; 131 T; 0 other;

alignment_scores:

Quality: 912.00 Length: 178
 Ratio: 5.124 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x AAS13480 ..

Align seg 1/1 to: AAS13480 from: 1 to: 543

```

29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAla 45
|||||
2 AGAGTTGCATGAGAAATCCAGTGTCCAGCATCTTGCCTGTGGC 51
|||||
45 AleuSerTYrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
|||||
52 CCCTCTCTACCTCTGGCCAGAGATCCACAGTCAACCTGGAGCCAAA 101
|||||
62 ysaSPThrYsaSPSerArgProLysLeuProGlnThrLysSerArgGly 78
|||||
102 AGGACACAAAGAGCTCTCGACCCAACTGCCAGACCCCTCTCAGAGGT 151
|||||
79 TTPGlyAspGlnLeuIleThrPthrGlnThrTyrgluGlnAlaLeuTyrl 95
|||||
152 TGGGGTGACCACTCATCTGACTCAGACATATGAAGAACCTCATATATA 201
|||||
95 sserLysThrSerAsnLysProLeuMetIleIleHisLeuAspGluC 112
|||||
202 ATCCAAAGACAGCAACAACTTGATGATTAATTCACCTTGATGATGAGT 251
|||||
112 ySPProHisSerGlnAlaLeuLysValPheAlaGluAsnLysGluIle 128
|||||
252 GCCCAGACAGTCAAGCTTTAAAGAAAGTGTTCGTAATAAAGAAATC 301
|||||
129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgluThr 145
|||||
302 CAGAAATTGGCAGACAGAGTTGTCTCCTCAATCTGGTTATGAAACAC 351
|||||
145 rAspLysHisLeuSerProAspGlyIleTyValProArgIleMetPhe 162
|||||
352 TGACAAACACCTTCTCTGATGAGCCAGTATGTCCTCCAGATATATGTT 401
|||||
162 AlaSPProSerLeuThrValArgAlaAspIleThrGlyArgTySerAsn 178
|||||
402 TTACCCCATCTCGACAGTTAGAGCCGATATCACTGGAAATATTCAAC 451
|||||
179 ArgLeuTyAlaTyrgluProAlaAspThrAlaLeuLeuAspAsnMe 195
|||||
452 CGTCTCATCTTACGAACTCGCAGATACAGTCTGTGCTGCAACAT 501
|||||
195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
502 GAAGAAAGCTCTCAAGTGTCTGAAGACTGAATTG 535
|||||

```

seq_name: /SIDS1/gcgsdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV29048

seq_documentation_block:

ID AAV29048 standard; CDNA; 525 BP.
 AC AAV29048;
 XX
 DT 21-AUG-1998 (first entry)

XX
 DE Open reading frame human protein comprising secretory signal 9.
 XX
 KW Human protein; secretory signal; nutritional source; cytokine;
 KW immunoty;haematopoesis; activin; inhibin; tumour; chemotactic;
 KW chemokinetic; thrombolytic; anti-inflammatory; inhibition;
 KW stomach cancer cell; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1..525
 FT CDS /tag=a
 FT /product="human protein comprising secretory
 FT signal"

WO9811217-A2.

19-MAR-1998.

12-SEP-1997; 97WO-JP03239.

13-SEP-1996; 96JP-0243060.

(PROT-) PROTEGENE INC.
 (SAGA) SAGAMI CHEM RES CENTRE.

Kato S, Kobayashi M, Sekine S, Yamaguchi T;

WPI; 1998-207380/18.

P-PSDB: AAM37872.

Human proteins with secretory signal sequences - used to treat
 PT immune deficiencies, infections, tumours, and haematopoietic
 PT disorders, etc.

Claim 3; Pages 88; 131pp; English.

XX This is the nucleotide sequence of the open reading frame of a novel
 CC human protein comprising a secretory signal (AAV29047), isolated from
 CC stomach cancer cells. Its proteins can be used as nutritional sources
 CC or supplements. The proteins may also have cytokine functions,
 CC immune modulating functions, haematopoesis regulating activity,
 CC activin/inhibin regulating activity, chemotactic/chemokinetic activity,
 CC haemostatic and thrombolytic activity, receptor/ligand activity,
 CC anti-inflammatory activity, tumour inhibition activity.
 XX

SQ Sequence 525 BP; 162 A; 131 C; 103 G; 129 T; 0 other;

alignment_scores:

Quality: 899.00 Length: 175
 Ratio: 5.137 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x AAV29048 ..

Align seg 1/1 to: AAV29048 from: 1 to: 525

```

32 MetGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerTy 48
|||||
1 ATGGAGAAATTCACAGTGCAGCATCTTCTCTCTGCGCCCTCTCTCTA 50
|||||
48 rThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThrL 65
|||||
51 CACTCTGGCCAGAGATACACAGTCAAACTGGAGCCAAAAGACACAA 100
|||||
65 ysaSPSerArgProLysLeuProGlnThrLysSerArgGlyTTPGlyASP 81
|||||
101 AGGACTCTCGACCCCAAACTGCCAGACCTCTCCAGAGTGGGGGAGAC 150
|||||
82 GlnLeuIleThrPthrGlnThrTyrgluGlnAlaLeuTyrlYserLysTh 98
|||||

```

```

|||||
151 CAACTCATCTGAGACTCAGACATATGAAGAAGCTCTATATAATCCAGAC 200
98 rSeAsnLysProLeuMetIleIeHsHsLeuAspGluCysProHss 115
201 AAGCAACAACCCCTGTGATTTATTTCTCATCTTGATGAGTGCACCAACA 250
115 eRgInAlaLeuLysValPheAlaGluAsnLysGluIleGlnLysLeu 131
251 GTCAGAGTTTAAAGAAAGTGTTCCTGAATAATAAGAAATCCAGAAATTG 300
132 AlAGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
301 GCAGAGCAGTTTGCTCTCCCAATCTGTTATGAACAACACTGCAACAACA 350
148 sLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 165
351 CCTTCTCCGATGCGCAGATATGCCAGGATATGTTGTTTCAACCAT 400
165 eRLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyr 181
401 CTCGACAGATTAGAGCCGATATCATCGAAGATATTCAAACCGTCTCTAT 450
182 AlATyrGluProAlaAspThrAlaLeuLeuAsnLysMetLysLysAl 198
451 GCTTACGAACTGCAGATACAGCTCTGTGCTGACACAACATGAAGAAAGC 500
198 aLeuLysLeuLeuLysThrGluLeu 206
501 TCTCAAGTGTCTGAAGACTGAATTG 525

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAZ24578

seq_documentation_block:

ID AAZ24578 standard: cDNA; 793 BP.

AAZ24578:

07-DEC-1999 (first entry)

Human lung tumor associated polynucleotide.

Human: lung tumor; lung cancer; T cell stimulation; ss.

Homo sapiens.

MO9947674-A2.

23-SEP-1999.

17-MAR-1999; 99WO-US05798.

18-MAR-1998; 98US-0040802.

18-MAR-1998; 98US-0040984.

27-JUL-1998; 98US-0123912.

27-JUL-1998; 98US-0123933.

(CORI-) CORIXA CORP.

Reed SG, Wang T;

WPI; 1999-571839/48.

New isolated lung tumor polynucleotides, used to develop products for the treatment, prevention and monitoring the progression of lung cancer

Claim 1: Page 90; 148pp; English.

The invention provides isolated human lung tumor nucleic acids and polypeptides. The polypeptides can be used for the treatment of lung cancer. The polypeptides and polynucleotides can be used to stimulate T cells or antigen presenting cells for use in the treatment of lung

CC cancer. The polypeptides and monoclonal antibodies specific for the
 CC polypeptides can also be used to inhibit the development of lung cancer.
 CC Agents which bind the polypeptides can be used for detecting lung cancer
 CC and for monitoring the progression of lung cancer.
 XX
 SQ Sequence 793 BP; 246 A; 172 C; 148 G; 211 T; 16 other;

alignment_scores:

Quality: 856.50 Length: 190
 Ratio: 4.706 Gaps: 1

Percent Similarity: 95.789 Percent Identity: 93.684

alignment_block:

US-09-674-266a-181 x AAZ24578 ..

Align seg 1/1 to: AAZ24578 from: 1 to: 793

```

14 HisProSerArgArgLeuThrGlnGlyArgTyrValArgLysSerArgVa 30
2 CATCTAGCCGCGCAGTCACACAAAGCGAGGTGGTGAAGAAATCCAGACT 51
30 lAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuS 47
52 TGCATGAGAAATTCAGATGTCAGCATTCCTGCTGCGCCCTCT 101
47 eRtYrThrLeuAlaArgAspThrValLysProGlyAlaLysLysAsp 63
102 CCTACACTGCGCCAGATACCAAGTCAACCACTGAGCCCAAAAGGAC 151
64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgLysTyrP 80
152 ACAAAGACTCTGACCCAACTGCCCCAGACCCTCTCCAGAGTTGGG 201
80 YAspGlnLeuIleTyrThrGlnThrTyrGluGluAlaLeuTyrLysSer 97
202 TGACCAACTCATCTGAGACTCAGACATATGAAGAAGCTTATATAATCA 251
97 yStrHsSerAsnLysProLeuMetIleIeHsHsLeuAspGluCysPro 113
252 AGACAAGCAACAAACCTTGATGATATTCATCACTGATGAGAGGCCCA 301
114 HisSerGlnAlaLeuLysValPheAlaGluAsnLysGluIleGlnLy 130
302 CACAGTCNAGCTTTAAAGAAAGTGTTCGTAATAATAAGAAATCCAGAA 351
130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrAspL 147
352 ATTGCGACAGCAGTTTGTCTCTCAATCTGTTATGAACAACACTGACA 401
147 ySHsLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
402 AACACCTTCTCCGATGCGCAGATATGCCAGGATATGTTGTTGAC 450
164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
451 CCATCTCTGACAGTTGAAGCCGATATCCTGGGAAGATATTCNAACCGCT 500
180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuAsnLysMetLysL 197
501 CTATGCTTACAAACT.GCAGATACGCTGTGTTGCTTGACACACT...GAAA 546
197 ySAlaLeuLysLeuLys 203
547 AAGCTCTCAAGTTGCTTNA 566

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA65817

seq_documentation_block:

ID AAC65817 standard: cDNA; 793 BP.

AA65817:

XX

DT 21-FEB-2001 (first entry)

DE Human lung cancer-associated CDNA LSF-S1-A-10A.

XX Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW vaccine; detection; ss.

OS Homo sapiens.

PN WO200061612-A2.

XX 19-OCT-2000.

PF 03-APR-2000; 2000WO-US08896.

PR 02-APR-1999; 99US-0285479.

PR 17-DEC-1999; 99US-0466396.

PR 30-DEC-1999; 99US-0476496.

PR 10-JAN-2000; 2000US-0480884.

PR 22-FEB-2000; 2000US-0510376.

XX (CORI-) CORIXA CORP.

PI Wang T, Fan L;

DR WPI: 2000-628399/60.

XX Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PS in a patient -

XX Claim 1a; Page 123; 261pp; English.

CC This invention describes a novel isolated polypeptide (I) which
CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC which have cytostatic activity. The polypeptides and polynucleotides are
CC used in compositions and vaccines to inhibit the development of cancer.
CC especially lung cancer, in a patient. Methods described in the invention
CC can be used to monitor the progression of a cancer by carrying out the
CC detection at subsequent time points and comparing the results from the
CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC are treated with P2, polynucleotides encoding P2 or antigen presenting
CC cells expressing P2 and then administered to the patient to inhibit
CC development of cancer.

SO Sequence 793 BP; 246 A; 172 C; 148 G; 211 T; 16 other:

alignment_scores: Quality: 856.50 Length: 190
 Ratio: 4.706 Gaps: 1
Percent Similarity: 95.789 Percent Identity: 93.684

alignment_block:
US-09-674-266A-181 x AAC65817 ..

Align seg 1/1 to: AAC65817 from: 1 to: 793

14 HIsProSerArgArgLeuThrGlnGlyArgTyrPValAlaArgLysSerArgVa 30
 |||||
2 CATCTAGCGCGCGACTACACAAAGCAGGTGGTGAGAAATCCAGAGT 51
30 lAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuS 47
 |||||
52 TGCCATGAGAAAAATTCAGTGTACGATTCCTGCTCCTGTGGCCCTCT 101
47 eTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysASP 63
 |||||
102 CTTACACTCTGGCCAGAGATACACAGTCAAACTGAGCCAAAAGAGAC 151
64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTyrG1 80
 |||||
152 ACAAGGACTCTCGACCAAACTGCCAGACCTCTCCAGAGTTGGGG 201

80 yAspGlnLeuIleTyrProThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
 |||||
202 TGACCAACTCATCTGGACTGACATATGAGAAAGCTTATATAATCA 251
97 ySThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluCyPro 113
252 AGACAAGCAACAACCTTGATGATTATTCATCAGTGGATGGGCCA 301
114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
302 CACAGTCNAGCTTTAAAGAAAGTGTGCTGAAATTAAGAAATCCAGAA 351
130 sleuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
352 ATTGCGAGAGCAGTTGTCTCTCAATCTGGTTATGAAACAACCTACA 401
147 ySHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
402 AACACCTTCTCCTGATGGCCAGTATGT.CCCAGGATTATGTTGTGAC 450
164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
451 CCATCTCTGACAGTTGAAGCCGATATCTGGAAAGATATTCNAACCGTCT 500
180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
501 CTATGCTTACAAACT.GCAGATAGCTCTGTGCTTGACACAT...GAAA 546
197 ySaLeuLeuLysLeuLeuLys 203
547 AAGCTCAAGTTGCTTAAA 566

OM of: US-09-674-266A-181 to: Issued_Patents_NA:* out_format : pfs
Date: Oct 8, 2002 10:03 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framed.p2n.model -DEV=x1h  
-O=/cgn2_1/USPTO.spool/US09674266/runat_08102002_093325_26183/app_query.fasta.1.271  
-DB=Issued_Patents_NA -Qfmt=fastap -SOFix=p2n.rn1  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=70.000  
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELDP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=dlscum2 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR_SCORE=pcpt -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFM=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09674266.ecgn1.1.46 -NCPu=6 -ICPu=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPY -WAIT -THREADS=1
```

Search information block:

Query: US-09-674-266A-181
Query length: 206
Database: Issued_Patents_NA:*
Database sequences: 383533
Search length: 122816752
Search time (sec): 46.580000

Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-1 +				991.00	2195.32	1.4e-114 875
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-247-155-61 +				961.00	2120.07	2.1e-110 1689
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-247-155-149 +				914.00	2021.63	6.4e-105 940
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-040-984-78 +				856.50	1894.61	7.7e-98 793
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-123-912-72 +				856.50	1894.61	7.7e-98 793
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-221-298-32 +				683.00	1513.40	1.3e-76 401
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-35 +				642.00	1421.83	1.7e-71 386
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-5 +				555.00	1218.86	3.3e-52 373
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-36 +				458.50	1075.46	3.1e-48 489
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-9 +				376.50	802.77	3.9e-38 506
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-10 +				253.00	555.46	1.8e-23 158
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-030-607-209 +				252.00	557.14	2.4e-23 159
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-439-313-209 +				252.00	557.14	2.4e-23 159
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-3 +				222.00	463.43	4.0e-18 1423
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-38 +				183.50	405.27	6.9e-15 136
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-19 +				158.00	338.43	3.7e-11 302
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-41 +				147.50	314.55	7.8e-10 310
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-11 +				129.00	272.99	1.6e-07 311
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-12 +				104.00	216.20	0.0002 329
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-13 +				104.00	216.20	0.0002 329
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-14 +				104.00	216.20	0.0002 329
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-15 +				101.00	209.43	0.0006 330
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-22 +				95.50	201.69	0.0015 225
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-534-638-1 +				88.50	140.52	3.87 9840
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-16 +				87.50	174.01	0.0528 505
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-18 +				82.00	166.93	0.1308 326
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-07-745-206A-14 -				79.50	136.95	6.12 2470
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-311-363-14 -				79.50	136.95	6.12 2470
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-07-745-206A-12 -				79.50	127.39	20.85 3467
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-311-363-12 -				79.50	127.39	20.85 3467
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-455-543A-8 -				79.50	124.12	31.71 7175
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-193-078B-8 -				79.50	124.12	31.71 7175
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-223-305C-8 -				79.50	124.12	31.71 7175
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-149-097D-8 -				79.50	124.12	31.71 7175
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-949-386-8 -				79.50	124.12	31.71 7175
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-450-562-8 -				79.50	124.12	31.71 7175
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-984-709A-8 -				79.50	124.12	31.71 7175
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-268-163-7 -				79.50	124.12	31.71 7177

/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-713-118-1 - 79.50 123.97 32.34 7266
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-452-007-1 - 79.50 123.97 32.34 7266
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-455-543A-7 - 79.50 123.81 33.00 7362
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-193-078B-7 - 79.50 123.81 33.00 7362
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-223-305C-7 - 79.50 123.81 33.00 7362

seq_name: /cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-916-576B-1

seq_documentation_block:

Sequence 1, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EMMER, REINHARD
APPLICANT: EMMER, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 875 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 71..595
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 131..595
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 71..130
US-08-916-576B-1

alignment_scores:

Quality: 991.00 Length: 194
Ratio: 5.135 Gaps: 0
Percent Similarity: 99.485 Percent Identity: 99.485

alignment_block:

US-09-674-266A-181 x US-08-916-576B-1 ..

Align seg 1/1 to: US-08-916-576B-1 from: 1 to: 875

```

13 ProHISProSerArgLeuThrGlnGlyArgTyrValArgLysSerAr 29
14 CCGATTCTTACGCCCGACTGACACAAAGCAGGTGGTGAGAAATCCAG 63
29 gValAlaMetGluLysIleProValSerLAlaPheLeuLeuValAlaL 46
64 AGTTGCCATGGAGAAATTCAGTGTACAGCATCTTGGCTCTGGGCC 113
46 eUSeTyrThrLeuAlaArgAspThrThrValLysProGluAlaLysLys 62
114 TCTCTCACTCTGGCCAGAGATACCACAGTCAAACTTGAGCCAAAAG 163
63 AspThrLysAspSerArgProLysLeuProGlnThrLysSerArgGlyTr 79
164 GACACAAGAGACTCTGCAGCAAGTCCAGACCCCTCTCCAGAGGTGG 213
79 pGlyAspGlnLeuIleTyrThrGlnThrTyrgluGluAlaLeuTyrLys 96
214 GGGTGACCAACTCATCTGAGCTCAGCATATGAGAGAGCTCTATATAAAT 263
96 eTyrThrSerAsnLysProLeuMetIleIleHisLysLeuSpGluLys 112
264 CCAAGACAAAGCAACAACCTTGATGATTAATTCATCCTTGATGAGTGC 313
113 ProHISSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleG 129
314 CCACACAGTCAAGCTTTAAAGAAAGTTGGCTGAATAATTAAGAAATCCA 363
129 nLysLeuAlaGluGlnPheValLeuAsnLeuValTyrGluThrThra 146
364 GAAATGGAGAGAGATTGTCTCTCATCTGTTATGAAACAACAG 413
146 sPlysHisLysSerProAspGlyGlnTyrValProArgIleMetPheVal 162
414 ACAAACACCTTTCTCTGATGGCCAGTATGTCGCCAGGATTAATGTTGTT 463
163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
464 GACCCATCTCTGACAGTTAGAGCCGATATCAGTGAAGATATTCAAATCG 513
179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
514 TGTCTATGCTTACGAACCTGCAGATACACGCTGTGTGCTTGACAACATGA 563
196 yAlaLysAlaLeuLysLeuLeuLysThrGluLeu 206
564 AGAAAGCTCTCAAGTTGCTGAAGAGACTGAATTC 595

seq_name: /cgn2_6/ptodata/1/ina/6b_COMB.seq:us-09-247-155-61
seq_documentation_block:
: Sequence 61. Application US/09247155A
: Patent No. 6312822
: GENERAL INFORMATION:
: APPLICANT: Dumas Milne Edwards, Jean-Baptiste
: APPLICANT: Duclef, Aymeric
: APPLICANT: Bouquellet, Lydie
: TITLE OF INVENTION: Complementary DNAs
: FILE REFERENCE: GENSET.021A
: CURRENT APPLICATION NUMBER: US/09/247,155A
: EARLIER FILING DATE: 1999-02-09
: EARLIER APPLICATION NUMBER: 60/074,121
: EARLIER FILING DATE: 1998-02-09
: EARLIER APPLICATION NUMBER: 60/081,563
: EARLIER FILING DATE: 1998-04-13
: EARLIER APPLICATION NUMBER: 60/096,116
: EARLIER FILING DATE: 1998-08-10
: EARLIER APPLICATION NUMBER: 60/099,273
: NUMBER OF SEQ ID NOS: 182
: SOFTWARE: Patent.pm
: SEQ ID NO 61
: LENGTH: 1689

```

```

: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 51..575
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 51..110
: OTHER INFORMATION: Von Heijne matrix
: OTHER INFORMATION: score 11.2
: OTHER INFORMATION: seq AFLUVALSTYLA/RD
: FEATURE:
: NAME/KEY: polyA_signal
: LOCATION: 1653..1658
: FEATURE:
: NAME/KEY: polyA_site
: LOCATION: 1674..1689
: US-09-247-155-61

```

```

alignment_scores:
    Quality: 961.00      Length: 195
    Ratio: 5.085        Gaps: 1
    Percent Similarity: 96.923    Percent Identity: 96.923

```

```

alignment_block:
US-09-674-266a-181 x US-09-247-155-61  ..

```

```

Align seg 1/1 to: US-09-247-155-61 from: 1 to: 1689

```

```

12 GlyProHISProSerArgLeuThrGlnGlyArgTyrValArgLysSe 28
9 GGAGCGCATCTTACGCCCGACTGACACAAAGC..... 41
28 rArgValAlaMetGluLysIleProValSerLAlaPheLeuLeuValA 45
42 .AGAGTGGCATGGAGAAATTCAGTGTACAGCATCTTGGCTCTGTGG 90
45 lAlaLeuSerTyrThrLeuAlaArgAspThrThrValLysProGluAlaLys 61
91 CCTCTCTTACACTCTGGCCAGAGATACCACAGTCAAACTTGAGCCAAA 140
62 LysAspThrLysAspSerArgProLysLeuProGlnThrLysSerArgL 78
141 AAGGACACAAGAGACTCTGCAGCAAACTGCCAGACCTCTCCAGAG 190
78 yTyrpGlyAspGlnLeuIleTyrThrGlnThrTyrgluGluAlaLeuTyrL 95
191 TTGGGGTGACCAACTCATCTGAGCTCAGCATATGAGAGAGCTCTATATA 240
95 ySeriLysThrSerAsnLysProLeuMetIleIleHisLysLeuSpGlu 111
241 AATCCAGACAAAGCAACAACCTTGATGATTAATTCATCCTTGATGAG 290
112 CysProHISSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluI 128
291 TCCCCACACAGTCAAGCTTTAAAGAAAGTTGGCTGAATAATTAAGAAAT 340
128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
341 CCAGAAATGGCAGAGCAAGTTGTCTCTCATATCTGTGTTTGAACAACA 390
145 hrAspLysHisLysSerProAspGlyGlnTyrValProArgIleMetPhe 161
391 CTGACAAACACCTTTCTCTGATGGCCAGTATGTCGCCAGGATTAATGTT 440
162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAr 178
441 GTTGACCATCTCTGACAGTTAGAGCCGATATCAGTGAAGATATTCAAA 490
178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
491 TGTCTCTATGCTTACGAACCTGCAGATACACGCTGTGTGCTTGACAACA 540

```

195 etlslsyalaleuylsleuleuylsthrGluleu 206
|||||
541 TGAAGAAAGCTCTCAAGTGTGCTGAAGCTGAATTG 575

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-247-155-149

seq_documentation_block:

; Sequence 149, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclet, Aymeric
; APPLICANT: Bouquellet, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 149
; LENGTH: 940
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 177..569
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 177..236
; OTHER INFORMATION: Von Heljne matrix
; OTHER INFORMATION: score 11.1999998092651
; OTHER INFORMATION: seq AFLIVALSITLA/RD
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 931..939
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 482
; OTHER INFORMATION: n=a, g, c or t
US-09-247-155-149

alignment_scores:
Quality: 914.00 Length: 207
Ratio: 4.811 Gaps: 2
Percent Similarity: 91.787 Percent Identity: 90.821

alignment_block:

US-09-674-266A-181 x US-09-247-155-149 ..

Align seg 1/1 to: US-09-247-155-149 from: 1 to: 940

1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerar 17
|||||
103 AGACTCGGCTGTGCTGACACACAGAGCTT.GGACCGCATCCTACCG 151
17 gATgLeuThrGlnGlyArGTrpValArGlySserArGValAlaMetGluL 34
|||||
152 CCGAGCTCACAAAGCC.....AGAGTTGCCATGGA 183
34 yslleProValSerAlaPheLeuLeuValAlaLeuSerTyrThrLeu 50
|||||
184 AAATTCAGTGTGACATCTCTGCTCTTGCGCCCTCTCTACACTCTG 233

51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAsp 67
|||||
234 GCCAGAGATACACAGTCAAACTGGAGCCAAAAGACACAAAGACTC 283
67 rArGProLysLeuProGlnThrLeuSerArGlyTrpGlyAspGluLeu 84
|||||
284 TCGACCCAAACTGCCACAGCCTTCACAGAGGTGGGTGACCAACTCA 333
84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
|||||
334 TGTGACACACACATATGAARARCTGTATWTAATCCARACACAGCAAC 383
101 LysProLeuMetIleIleHisLysLeuAspGluCysProHisSerGlnAl 117
|||||
384 AAACCTGTGATGATATTCATCTGATCGATGATGCCACACAGTCAAGC 433
117 alauLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
|||||
434 TTTAAAAAAAKTGTCTGCTGAAAAATAAARAATCCAGAAATTTGGCARNC 483
134 lnhPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
|||||
484 AGTTGTCCCTCAATCTGTTATGAACAACACTGACAAACACTTTCT 533
151 ProAspGlyGlnTyrVal.ProArGileMetPheValAspProSerLeu 167
|||||
534 CCGATGCGCCAAATATKTCGCCCMGAGATTATGTTGTTGACCCATCTCTGA 583
167 hrValArgAlaAspIleThrGlyArGTrpSerAsnArGLeuTyrAlaTyr 183
|||||
584 CAGTTAGAGCCGATATCATCATCGAAGATATCAAAACGCTCTATGCTTAC 633
184 GluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeu 200
|||||
634 GAACCTGCAGATACAGCTGTGCTTGACACATGAGAAAGAACTCTCA 683
200 slauLeuLysThrGluleu 206
|||||
684 GTTGCTCAACACTGAATTG 702

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-040-984-78

seq_documentation_block:

; Sequence 78, Application US/09040984
; Patent No. 6210883
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF LUNG CANCER
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,984
; FILING DATE: 18-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: MAKI, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.456
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900

TELEFAX: 206-282-6031
TELEX:
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-040-984-78

alignment_scores:
Quality: 856.50 Length: 190
Ratio: 4.706 Gaps: 1
Percent Similarity: 95.789 Percent Identity: 93.684

alignment_block:
US-09-674-266a-181 x US-09-040-984-78 ..

Align seg 1/1 to: US-09-040-984-78 From: 1 to: 793

```
14 HisProSerArgLeuThrGlnGlyArgTyrValArgLysSerArgVa 30
2 CATCTGAGCCGCGCAGTCACACAGCAGGTGGGTGAGAAATCCAGAGT 51
30 LAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuS 47
52 TGGCATGAGAAATTCAGAGTCAGCATCTCTCTGTGGCCCTCT 101
47 ertYrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
102 CCTCACTGCTGGCCAGATACCCAGCTCAAACTGGAGCCAAAAGGAC 151
64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgLysTyrPgl 80
152 ACAAGAGCTCTGACCCAAACTGCCAGACCTCTCCAGAGGTTGGGG 201
80 yAspGlnLeuIleTyrThrGlnThrLysGluAlaLeuTyrLysSerL 97
202 TGACCAACTCATCTGAGACTGACATATGAGAGAGCTCTATATTAATCCA 251
97 ystHrSerAsnLysProLeuMetIleIleHisLysLeuAspGluCysPro 113
252 AGACAGAGCAACAACTTGATGATATTCATCACTGGATGAGCTGCCCA 301
114 HisSerGlnAlaLeuLysLysValPheAlaGluLysLysGluIleGlnLys 130
302 CACAGTCAGCTTTAAAGAAAGTGTTCCTCAAAATTAAGAAATCCAGAA 351
130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
352 ATGGGAGAGAGAGTTTGTCTCTCAATCTGGTTATGAAACAACACTGACA 401
147 yshLysLeuSerProAspGlnLysIleTyrValProAlaGlyIleMerPheValAsp 163
402 AACACCTTTCCTCGATGGCCAGATATG. CCCAGATATATTTTGTGTAC 450
164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
451 CCATCTCTGAGAGTTGAAGCCGATATCTCTGGGAAGATATTCNAACGCTCT 500
180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuAspAsnMetLysL 197
501 CTATGCTTACAAACT. GCAGATACGCTCTGTGCTTGACACAT... GAAA 546
197 ysaLalaLeuLysLeuLys 203
547 AAGCTCTCAAGTTGCTNAAA 566
```

seq_name: /cgn2_6/prodata/1/lna/68_COMB.seq:us-09-123-912-78

seq_documentation_block:
; Sequence 78, Application US/09123912A

```
Patent No. 6312695
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Wang, Tongcong
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
FILE REFERENCE: 210121.455C1
CURRENT APPLICATION NUMBER: US/09/123,912A
PRIOR APPLICATION NUMBER: 1998-07-27
PRIOR FILING DATE: 1998-03-18
NUMBER OF SEQ ID NOS: 114
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 78
LENGTH: 793
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (309)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (492)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (563)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (657)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (660)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (703)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (708)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (710)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (711)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (732)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (740)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (748)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (758)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (762)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (765)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (787)
OTHER INFORMATION: Where n is a, c, g or t
US-09-123-912-78
```

alignment_scores:
Quality: 856.50 Length: 190
Ratio: 4.706 Gaps: 1
Percent Similarity: 95.789 Percent Identity: 93.684

alignment_block:
US-09-674-266a-181 x US-09-123-912-78 ..

Align seg 1/1 to: US-09-123-912-78 from: 1 to: 793

```

14 HisProSerArgArgLeuThrGlnGlyArgTyrPvalArgLysSerArgVa 30
   |||||||
2  CATCTTAGCGCGCGACTCACAGCAAGCAGGTGGTGAGAAATCCAGAGT 51
   |||||||
30 1AaMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuS 47
   |||||||
52 TGCACATGGAGAAATTCAGGTGCACATCTTGCTGCTGGGCCCTCT 101
   |||||||
47 eTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
   |||||||
102 CCAACACTCTGGCCAGAGATACCAAGCTCAAACTGAGGCCAAAAGAGAC 151
   |||||||
64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgLysTyrPgl 80
   |||||||
152 ACAAGGACTCTGCAGCCAAACTGCCAGACCTCTCCAGAGTTGGGG 201
   |||||||
80 yAspGlnLeuIleTyrThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
   |||||||
202 TGAACCACTCATCTGCAGCTCAGACATATGAAGAAGCTATATATAATCCA 251
   |||||||
97 yThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
   |||||||
252 AGCAAGGCAACAAACCTTGATGATATATTCATCTTGATGGATGGCCCA 301
   |||||||
114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnL 130
   |||||||
302 CACAGTCNMGCTTTAAAGAAAGTTGGTGGAATAATAAGAAATCCAGAA 351
   |||||||
130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
   |||||||
352 ATGGAGAGACAGTTGTCTCTCTCAATCGTTTATGAAGAAACAGTACA 401
   |||||||
147 yHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 153
   |||||||
402 AACACCTTTCTCTGATGGCCAGATATGT.CCCAGGATTTATGTTGTGAC 450
   |||||||
164 ProSerLeuThrValAlaArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
   |||||||
451 CCATCTCTGACAGTTGAAGCCGATATCCTGGGAAGATATTCAACGCTCT 500
   |||||||
180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
   |||||||
501 CTATGCTTCAAACT.GCAGATACGCTGTGTGCTTGACACAT...GAA 546
   |||||||
197 yAlaLeuLysLeuLeuLys 203
   |||||||
547 AAGCTCTCAAGTTGCTMAAA 566

```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-221-298-32

seq_documentation_block:

```

; Sequence 32, Application US/09221298
; Patent No. 6284241
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER
; FILE REFERENCE: 210121.471
; CURRENT APPLICATION NUMBER: US/09/221,298
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Human
US-09-221-298-32

```

alignment_scores:

Quality: 683.00 Length: 134
Ratio: 5.174 Gaps: 1
Percent Similarity: 98.507 Percent Identity: 98.507

alignment_block:
US-09-674-266a-181 x US-09-221-298-32 ..

Align seg 1/1 to: US-09-221-298-32 from: 1 to: 401

```

57 LysProGlyAlaLysLysAspThrLysAspSerArgProLysLeuProGln 73
   |||||||
2  AAACCTGGAGCCAAAAGAGACAAAGAGACTCTGCAGCCCAAACTGCCCA 51
   |||||||
73 nThrLeuSerArgGlyTyrPglLysProGlnLeuIleTyrThrGlnThrTyrG 90
   |||||||
52 GACCTCTCCAGAGTTGGGTGACCAACTCATCTGACACAGATATATG 101
   |||||||
90 LuGluAlaLeuTyrLysSerLysThrSerAsnLysProLeuMetIleIle 106
   |||||||
102 AACAGCTCTATATATAATCCAGACAAAGCAAAACCTTGATGATATT 151
   |||||||
107 HisHisLeuAspGluCysProHisSerGlnAlaLeuLysLysValPheAl 123
   |||||||
152 CATCACTTGGGTGAGTGGCCACACAGTCAAGCTTTAAAGAAATGTTTGC 201
   |||||||
123 agLysLysGluIleGlnLysLeuAlaGluGlnPheValLeuLeuAsnL 140
   |||||||
202 TGAATAATAAGAAATCCAGAAATTTGGCAGAGCAGTTGTCTCTCAATC 251
   |||||||
140 euValTyrGluThrThrAspLysHisLeuSerProAspGlyGlnTyrVal 156
   |||||||
252 TGGTTATGAAACACTGACAAACACCTTTCTCTGATGGCCAGTATGTC 301
   |||||||
157 ProArgIleMetPheValAspProSerLeuThrValAlaArgAla.AspIleT 173
   |||||||
302 CCAGAGTTATGTTGTTGACCATCTCTGACAGTTAGAGCCCGGATATCA 351
   |||||||
173 hArgLysTyrSerAsnArgLeuTyrAlaTyrGluProAlaAspThrAla 189
   |||||||
352 CTGGAAGATATTCAAAACCGTCTCTATGCTTACGAACTGCGAGATACAGCT 401
   |||||||

```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-35

seq_documentation_block:

```

; Sequence 35, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347

```

```

FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEEFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-916-576B-35

```

```

alignment_scores:
Quality: 642.00      Length: 128
Ratio: 5.016         Gaps: 0
Percent Similarity: 100.000    Percent Identity: 97.656

```

alignment_block:

US-09-674-266A-181 x US-08-916-576B-35 ..

Align seg 1/1 to: US-08-916-576B-35 from: 1 to: 386

```

61 LysLysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArg 77
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2  CGACGAGCACAAGAGACTCTGACCAACTGCCAGACCCCTCTCAG 51
77 GGLYTPSLYASPGINLEUILETRPTHTGINTHTYRGLUALALEUT 94
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
52 AGGTGGGGTGACAA.CTCATCTGACCTCAGACATATGAAGAGCTCAT 100
94 YRLYSERTLYSERTSERANLYSPROLEUMETILEIHISHISLEUASP 110
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
101 ATAAATCCAAAGACAGCAAAACCTTGATGATATTCATCATCTTGAT 150
111 GLUCYSPROHISSEKGLALALEUYSLYSVALPHEALAGLUAASLYGL 127
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
151 GAGTCCCAACAGTCACAGCTTAAAGAAAGTTGCTGAAATAAAGA 200
127 uileglnlyseuallagluinphevalleuueasneuvaltyrglut 144
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
201 AATCCAGAAATTGGCAGAGAGATTGTCTCTCAATCGTTATGAAA 250.
144 hrthasplyshisleuserproaspglylntrvalproargllemet 160
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
251 CAACGTGACAAACACCTTCTCTGATGGCCAGTATGCCCCAGATTTATG 300
161 PheValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSe 177
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
301 TTGTGTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAGATATTC 350
177 rasnargleutyralatyrgluuproalaaspthr 188
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
351 AAATGCTCTATGCTTAGGAACCTGCAGATACA 384

```

seq_name: /cgn2_6/ptodata1/lna/6B_COMB.seq:US-08-916-576B-5

seq_documentation_block:

```

Sequence 5, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DILON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:

```

```

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEEFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 732 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 49..546
NAME/KEY: mat_peptide
LOCATION: 118..546
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 49..117
US-08-916-576B-5

```

```

alignment_scores:
Quality: 555.00      Length: 167
Ratio: 3.801         Gaps: 4
Percent Similarity: 87.425    Percent Identity: 64.671

```

alignment_block:

US-09-674-266A-181 x US-08-916-576B-5 ..

Align seg 1/1 to: US-08-916-576B-5 from: 1 to: 732

```

41 LeuLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspThrThrVally 57
   :::::::::::::::::::::::::::::
79 CTCTTACTCGTCACAGTTCTTCCAACTTGCC.....ATTGCAATAAA 122
57 sprolylvalalyslyaspthrlyaspsersargprolysleuprogint 74
   :::::::::::::::::::::::::::::
123 A.....AAGGAAAGAGGCT.....CCTCAGA 145
74 hrleuserarglytyrpslyaspglnleuiletrpthtglnthrtyslu 90
   :::::::::::::::::::::::::::::
146 CACTCTCAAGAGATGGGAGATGACATCACTTGCTGACCAAACTTATGAA 195
91 GluAlaLeuTyrLysSerLysThrSerAnlySPROLEUMETILEIH 107
   :::::::::::::::::::::
196 GAAGTCTCTTTATGCTCAAAAAGTAAGAAGCCATTAAATGTTATGCA 245
107 shisleuaspglucysprohisserglnalaleuulslysvalphealag 124
   :::::::::::::::::::::
246 TCACCTGGAGATGTGCAATCACTCAAGCACTAAAGAAAGTATTTGCC 295

```

```

124 luasnlysgluileglnlyslleualaglu... glnphevalleuleuasn 139
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 AAAATGAAGAAATACAGAAATGCTCAGAAATAGTTCATCATGCTAAAC 345
140 LeuValTyrGluThrThrAspLysHisLeuSerProAspGlyGlnTyrVa 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
346 CTATGTCAGTAAACCACTGATAGAAATTTATCATCTGATGGCAATATG 395
156 lProaGlyIleMetPheValAspProSerIleuThrValArgAlaAspIleT 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
396 GCCTAAACATGATGTTGTAGACCCCTTTAAACAGTTAGAGCTGACATAG 445
173 hrclyArGtyrSerAsnArgLeuTyrAlaTyrGluProAlaAspThrAla 189
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
446 CTGGAAGATACCTTAACAGATTGTACATATGAGCTCGGATTTAAACC 495
190 LeuLeuLeuAspAsnMetLysLysAlaLeuLysLeuLeuLysThrGluLe 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
496 CTATGATGTAAGAAACATGAAGAAAGCATTAAGACTTATTCAGTCAGAGCT 545
206 u 206
546 A 546
seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-36
seq_documentation_block:
; Sequence 36, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P. L. L. C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 373 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-916-576B-36

```

```

alignment_scores:
  Quality: 487.50      Length: 110
  Ratio: 4.643        Gaps: 3
  Percent Similarity: 95.455  Percent Identity: 91.818

```

```
alignment_block:
US-09-674-266a-181 x US-08-916-576B-36 ..
```

```
Align seg 1/1 to: US-08-916-576B-36 from: 1 to: 373
```

```

16 SerArgArgLeuThrGlnGlyArgValArgLysSerArgValAlaIle 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 AGCCGCCGACTCAGACAGGAGGAGTGGTGAGGAAATCCAGAGTTCAT 50
32 tGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerTyrT 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 GGAGAAATTCAGTGTGACGATTCCTGCTGCTGCTGCTGCTGCTGCTGCT 100
49 hrclyAlaArgAspThrThrValLysProGlyValAlaLysLysAspThrLys 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 CTCTGGCCAGAGATACCAAGTACGATTAACCTGGAGCCAAAGGACACAAAG 150
66 AspSerArgProLysLeuProGlnThrLeuSerArgLysTyrGlyAspG 82
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 GACTCTCGACTAAACTGCCCGACACCTCTCCAGAGTTGGGTGACCA 200
82 nLeuIleTyrThrGlnThrTyrGluAlaLeuTyrLysSerLysThr. 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 ACTCATCTGACTCAGACATATGAGAGCTCTATTAATCCAAAGACAN 250
99 SerAsnLysProLeuMetIleIleHisLysLeuAspGluCysProHis 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 GGCACAAACCTTGATGATTATTCATCACTTGGATGAGTGCCACACA 300
115 eArgAlaLeu...LysLysValPhe 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 GTTCANGTTTAAAGCAAGTTT 326
seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-9
seq_documentation_block:
; Sequence 9, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P. L. L. C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 489 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-916-576B-9

```

```

alignment_scores:
  Quality: 458.50      Length: 166
  Ratio: 3.728        Gaps: 9
Percent Similarity: 74.096   Percent Identity: 70.482

```

```

alignment_block:
US-09-674-266A-181 x US-08-916-576B-9  ..

```

```

Align seg 1/1 to: US-08-916-576B-9 from: 1 to: 489

```

```

11 SerGIyPpRoHISpRoSerArgrArgLeuthrgInglYArgrTPvAlArgLy 27
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4  TCGGCAGAGATCCrTAAGCCCGACTCAGACACAGCC..... 39
27 sSerArgrValAlaMetGIuLySIaPrOvAlSerAlaPhetLeuLeuY 44
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
40 ...AGACTTGCATGAGAAAATTCAGTGTCACTTCCTGCTCCCTTG 85
44 aAlaLeuSerTyrrThrLeuAlaArgrArgrThrValLySProGIyAla 60
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
86 TGGCCCTCTCTACACTCTGGCCAGATACCAAGTCACACCTGAGACC 135
61 LysLysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerAr 77
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
136 AAAAAGGACACAAAGACTTTCGACCAAACTGCCAGACCTCTCCAG 185
77 gGIyrrpGIyAspGIuLeu.IleTrrpHrgInThrTyrgIuGIuAla.Le 93
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
186 AGCTTGGGGTGCACACTTCATCTGACTCAGACATATGAGAAAGCTCTT 235
93 uTyr.LysSerLysThrSerAnLysProLeuMetIleIleHisHisLeu 109
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
236 ATATNAATTCACAGACAGCAACAAACCTTGATGATATTCATCATCTTG 285
110 AspGIuCySPRoHISerGIuAla..LeuLySlySValPheAlaGIuAsn 125
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
286 GGTGAGTCCACACAGCTTCAAGTTTAAAGAAAGTGTTCG..... 329
126 LysGIuIleGIuLysLeu.....AlaGIuGIuPheVal 136
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
330 .....TGGAATTAANGAANTCCGNAATTGGGNNAGACAGATTGT 370
136 l.LeuLeuAsnLeu.ValTyrgIuThrThrAspLysHisLeuSerProAs 152
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
371 NCCCNNTAAATTTGGTGGTTTGGAAACATGGGCAACACCTTTTGTG 420
152 pGIyInTyrrValProArgrIleMetPheValAspPro 164
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
421 NTGNCNATTTTTTCCCGGNTTTTTTTTTTTTGGCCCC 457

```

```

seq_name: /cgn2_6/prodata/1/lna/6B_COMB.seq:US-08-916-576B-10

```

```

seq_documentation_block:
; Sequence 10, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.

```

```

; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 506 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-916-576B-10

```

```

alignment_scores:
  Quality: 376.50      Length: 90
  Ratio: 4.482        Gaps: 3
Percent Similarity: 93.333   Percent Identity: 87.778

```

```

alignment_block:
US-09-674-266A-181 x US-08-916-576B-10  ..

```

```

Align seg 1/1 to: US-08-916-576B-10 from: 1 to: 506

```

```

97 LysThrSerAnLysProLeuMetIleIleHisHisLeuAspGIuCySPr 113
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
14 AAGACAGACACAAACCTTGATGATATTCATCATNTTGGATGAGTGGCC 63
113 cHisSerGIuAlaLeuLysLysValPheAlaGIuAsnLysGIuIleGln 130
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
64 ACACAGTCAGCTTAAAGAAAGTGTTCGAAATAAGAAACCAACGAGA 113
130 ySIeuAlaGIuGIuPheValLeuLeuAsnLeuValTyrgIuThrThrAs 146
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
114 AATTGGCAGAGAGATTGTCTCTCTCAATCTGTTATGAAACAACTGGA 163
146 pLysHisLeuSerProAspGIyGlnTyrrValProArgrIleMetPheVal 163
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
164 CAACACCTTCTCTCTGATGGCAGTATGTNCCAGGCTTATGTTGTG 213
163 sProSerLeu.ThrValArgrAlaAspIleThrgIyArG...TyrSerAs 178
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
214 ACCCATNTCTGACAGTACAGCCGATATCACTGGAGAGATATTCAAAC 263
178 nArgLeuTyrrAlaTyrr 183
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
264 CGTNNCTTATGCTTAC 279

```

```

seq_name: /cgn2_6/prodata/1/lna/6B_COMB.seq:US-08-916-576B-37

```

```
seq_documentation_block:
; Sequence 37, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-916-576B-37

alignment_scores:
Quality: 253.00 Length: 51
Ratio: 4.961 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x US-08-916-576B-37 ..
Align seg 1/1 to: US-08-916-576B-37 from: 1 to: 158

15 ProSerArgLeuThrGlnGlyArgTIPValAlaGlySerArgValAla 31
|||||
6 CTRAGCGCGCGACTCACACAGGAGGTGGTGAAGAAATCCAGAGTGC 55
|||||
31 aMetGluValIleProValSerAlaPheLeuLeuValAlaLeuSer 48
|||||
56 CATGGAGAAATTCACAGTGTACAGATTCCTGCTGCTGGCCCTCTCCT 105
|||||
48 yThrLeuAlaArgAspThrThrValIleProGlyAlaIleValIleVal 64
|||||
106 ACACTGTGGCCAGATACACAGTCAAACTGAGGCCAAAGGACACACA 155
|||||
65 Lys 65
|||||
156 AAG 158
```

```
seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-030-607-209
seq_documentation_block:
; Sequence 209, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: DILLON, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-030-607-209

alignment_scores:
Quality: 252.00 Length: 52
Ratio: 4.846 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x US-09-030-607-209 ..
Align seg 1/1 to: US-09-030-607-209 from: 1 to: 159

25 ValArgIleSerArgValAlaMetGluValIleProValSerAlaPhe 41
|||||
3 GTGAGGAATCCAGAGTGTGCATGGAGAAATTCACAGTGTCTCT 52
|||||
41 uLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspThrThrValIle 58
|||||
53 GCTCTGTGTGGCCCTCTCTACACTGTGGCCAGATACACAGTCAAC 102
|||||
58 roGlyAlaIleValIleValIleValIleValIleValIleValIle 74
|||||
103 CTGGAGCCCAAGAGACACAAAGGACTCTGACCCCAAACTGCCCAAGC 152
|||||
75 LeuSer 76
|||||
153 CTCCTCC 158

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-439-313-209
seq_documentation_block:
; Sequence 209, Application US/09439313
```

```

; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer E.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqun
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 209
; LENGTH: 159
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-439-313-209

```

```

alignment_scores:
  Quality: 252.00      Length: 52
  Ratio: 4.846        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:
US-09-674-266A-181 x US-09-439-313-209  ..

```

```

Align seg 1/1  to: US-09-439-313-209  from: 1  to: 159

```

```

25 ValArgLysSerArgValAlaMetGluLysIleProValSerAlaPheLe 41
|||||
3 GTGAGGAATCCAGAGTGCATCGAGAAATTCAGTCAGCATCTCTT 52
|||||
41 uLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspThrValLysP 58
|||||
53 GCTCCTTGCGCCCTCTCTACACTCTGCCAGATACCAAGTCAAGTCAAC 102
|||||
58 roGlyAlaLysLysAspThrLysAspSerArgProLysLeuProGlnThr 74
|||||
103 CTGAGGCCAAAGACGACCAAGACGCTCTGACCCAACTGCCCGACGACC 152
|||||
75 LeuSer 76
|||||
153 CTCCTCC 158

```

```

seq_name: /cgn2_6/ptodata/1/lna/6b_comb.seq:us-08-916-576B-3

```

```

seq_documentation_block:
; Sequence 3, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: XU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, RETINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1468.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 88..603
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 157..603
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 88..156
US-08-916-576B-3

```

```

alignment_scores:
  Quality: 222.00      Length: 158
  Ratio: 2.242        Gaps: 3
  Percent Similarity: 62.658  Percent Identity: 34.177

```

```

alignment_block:
US-09-674-266A-181 x US-08-916-576B-3  ..

```

```

Align seg 1/1  to: US-08-916-576B-3  from: 1  to: 1423

```

```

66 AspSerArgProLysLeuProGlnThr.....LeuSerArgLysTrpG 74
|||||
91 GAGACGGCGGCTCTGCTGGGCGCACCTGTTGCTGGGCTTCAGTTTCT 140
|||||
75 .....LeuSerArgLysTrpG 80
|||||
141 GCTCTCTGTCATCTCTCTGATGACATATGCGCTTGAAAGGATTTCG 190
|||||
80 LysArgLysLeuIleThrThrLysLysLysLysLysLysLysLysSer 96
|||||
191 GAGATCATATTCATTG...AGGACACTGGAAGATGGAGAGAAAGACCA 237
|||||
97 LysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluCysPr 113
|||||
238 GCTGCACATGACATGCCCCCTGATGTGATTTTCATTAATCTGTGTGG 287
|||||
113 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGln 130
|||||
288 AGCTTCAAGCTCTTAAGCCCAATTGTCAGAAATCTGAGAAATTTCAG 337
|||||
130 LysLeuAlaGluGlnPheValLeuLeuAsnLeu.....ValTyrGluThr 144
|||||
338 AACTCTCCCATATTTTGTATGTGAATCTTGAGAGTGAAGAGGACCC 387
|||||
145 ThrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPh 161
|||||

```



```

|||||
62 CAGAGTTGCCATGAGAAATTCAGTGCACATCTGCTGTGTG 111
45 lalseuSerThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
|||||
112 CCTCTCCCTACACTCTGGCCAGAGATCCACAGTCAAACTGGAGCCAA 161
62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgL 78
|||||
162 AAGGACAAAGAGACTCTGACCCAAACGCCCAACCCCTCCAGAG 211
78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrgluAlaLeuTyrL 95
|||||
212 TTGGGGTGCACCAATCATCTGACTCAGACATATGAAAGACTCTATATA 261
95 ySerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
|||||
262 AATCCAAAGACAAACAAACCCCTGATGATATTCATCATCTGGATGAG 311
112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluL 128
|||||
312 TGCCACACAGTCAGCTTTAAGAAAGTGTGCTGGAATAATGAAGAAAT 361
128 eGlnLysLeuAlaGluGlnPheValLeuLysLeuValTyrgluThrT 145
|||||
362 CCAGAAATGGCAGAGACAGTGTCTCCTCAATCTGGTTATGAAACAA 411
145 hAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
|||||
412 CTGACAAACACCTTCTCTCTGATGGCCAGTATGTCGCCAGATATATGT 461
162 ValAspProSerLeuThrValAlaArgAlaAspIleArgTyrSerAs 178
|||||
462 GTTGACCCATCTCTGACAGTTAGAGCCGATATCAGTGAAGATATTCAA 511
178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAs 195
|||||
512 TCGTCTCTATGCTTACGAACCTGCAATACACCTGTTGCTTGACACAA 561
195 eLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
562 TGAAGAAACCTCTCAAGTTGCTGAAGACCTGAATTTG 596
seq_name: gb_esc1:AM582256
seq_documentation_block:
LOCUS AM582256 689 bp mRNA linear EST 16-MAR-2000
DEFINITION OVA-ST0212-120100-075-e10 ST0212 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM582256
VERSION AM582256.1 GI:7257305
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 689)
HCGP http://www.ludwig.org.br/ORESTES.
The FAPESP/LICR Human Cancer Genome Project
UNPUBLISHED (1999)
CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV4&lt2=QV4-ST0212-
120100-075-e10&t3=2000-01-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 687.

```

```

FEATURES
    source
        Location/Qualifiers
            1..689
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_id="ST0212"
                /dev_stage="Adult"
                /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                Site_2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
BASE COUNT      212 a      172 c      143 g      162 t
ORIGIN
alignment_scores:
    Quality: 1008.00      Length: 195
    Ratio: 5.169      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-674-266a-181 x AM582256
Align seg 1/1 to: AM582256 from: 1 to: 689
12 GYPRHHISPROSEARARGRLEUHRGNGLYARGTRPVALARGLYSE 28
|||||
25 GGACCGCATCTTAGCCGCCGACTCACACAGCAGGTGGTAGAAGAAATC 74
28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValA 45
|||||
75 CAGAGTGGCATGAGAAATTCAGTGCACATTCCTGCTCCTGTGTG 124
45 lalseuSerThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
|||||
125 CCTCTCCCTACACTCTGGCCAGAGATACCAAGTCAAACTGGAGCCAA 174
62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgL 78
|||||
175 AAGGACAAAGAGACTCTGACCCAAACCTGCCACACCTCTCCAGAG 224
78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrgluAlaLeuTyrL 95
|||||
225 TTGGGGTGCACCAATCATCTGACATCAGACATATGAAGACTATATA 274
95 ySerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
|||||
275 AATCCAAAGACAAACAAACCCCTGATGATATTCATCATCTGGATGAG 324
112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluL 128
|||||
325 TGCCACACAGTCAGCTTTAAGAAAGTGTGCTGGAATAATGAAGAAAT 374
128 eGlnLysLeuAlaGluGlnPheValLeuLysLeuValTyrgluThrT 145
|||||
375 CCAGAAATGGCAGAGCACTTGTCTCTCAATCGGTTATGAAACAA 424
145 hAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
|||||
425 CTGACAAACACCTTCTCTGATGGCCAGTATGTCGCCAGATATATGTT 474
162 ValAspProSerLeuThrValAlaArgAlaAspIleArgTyrSerAs 178
|||||
475 GTTGACCCATCTCTGACAGTTAGAGCCGATATCAGTGAAGATATTCAA 524
178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAs 195
|||||
525 TCGTCTCTATGCTTACGAACCTGCAATACACCTGCTGCTTGACACAA 574
195 eLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||

```


/note="Organ: prostate; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."

BASE COUNT 234 a 174 c 158 g 183 t 1 others
ORIGIN

alignment_scores:
Quality: 983.00 Length: 191
Ratio: 5.147 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x BG179160 ..

Align seg 1/1 to: BG179160 from: 1 to: 750

```

15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
|||||
15 CCTAGCCGCCGACATCACACAGGCGGTGGAGAAATCCAGAGTTGC 64
|||||
31 aMeGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerT 48
|||||
65 CATGGAGAAATTCAGTGCAGATTCTCTCTGCGCCCTCTCTCT 114
|||||
48 YrThrLeuAlaArgAspThrThrValLysProGlyAlaLysAspThr 64
|||||
115 ACACCTCTGCGCAGAGATACACAGCTCAAACTGGAGCCAAAAGACACA 164
|||||
65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
|||||
165 AAGGACTCTGCACCAAACTGCCAGACCTCTCCAGAGTTGGGGTGA 214
|||||
81 pGlnLeuIleThrPheGlnThrTrpGluAlaLeuTyrlsSerLysT 98
|||||
215 CCAACTCTCTGAGACTCAGACATGAGAAAGCTTATATTAATCCAGA 264
|||||
265 CAGGACAAACCCCTTGATGATTTATTCATCTGGATGAGTGCACACAC 314
|||||
115 SerGlnAlaLeuLysValPheAlaGluAsnLysGluIleGlnLysLe 131
|||||
315 AGTCAAGCTTAAAGAAAGTGTGGCTGAATAAAGAAATCCAGAAAT 364
|||||
131 uAlaGluInPheValLeuLeuAsnLeuValTyrlGluThrThrAspLysH 148
|||||
365 GGCAGAGCAGTTTGTCTCTCAATCTGTTATGAAACAACTGCACAAAC 414
|||||
148 IsLeuSerProAspGlyGlnTrpValProArgIleMetPheValAspPro 164
|||||
415 ACCCTTCTCTGATGAGCCAGATATGCCCCAGAGATATATTTGTTGACCCA 464
|||||
165 SerLeuThrValArgAlaAspIleThrGlyArgTyrlSerAsnArgLeuTy 181
|||||
465 TCTGTGACAGATTAGAGCCGATATACCTGGAAGATATCAAAATCGCTCTA 514
|||||
181 rAlaTyrluProAlaAspThrAlaLeuLeuLeuAsnMetLysLysA 198
|||||
515 TGCCTACACACCTGCGATACAGCTCTGTTGTCGTAACAACATGAGAAAG 564
|||||
198 IaLeuLeuLeuLeuLysThrGlu 205
|||||
565 CTCTCAAGTTGCTGAGAGACTGAA 587

```

seq_name: gb_esc12:BE748141

seq_documentation_block:

LOCUS BE748141 689 bp mRNA linear EST 15-SEP-2000
DEFINITION 601571509p1 NIH-MGC_55 Homo sapiens cDNA clone IMAGE:3838347 5',
mRNA sequence.

ACCESSION BE748141 GI:10162133
VERSION BE748141.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 689)
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
plate: LICM525 row: a column: 04
High quality sequence stop: 643.
Location/Qualifiers

FEATURES

source
1..689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH-MGC_55"
/tissue_type="from acute myelogenous leukemia"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgctcgcc); Site_2: SfiI (ggccatagggc
)'; Double-stranded cDNA was prepared from cell line RNA,
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATTCTAGAGCGGCGGCGCCACATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT 209 a 170 c 148 g 162 t
ORIGIN

alignment_scores:
Quality: 980.00 Length: 191
Ratio: 5.131 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.476

alignment_block:
US-09-674-266A-181 x BE748141 ..

Align seg 1/1 to: BE748141 from: 1 to: 689

```

15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
|||||
3 CCTAGCCGCCGACATCACACAGGCGGTGGAGAAATCCAGAGTTGC 52
|||||
31 aMeGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerT 48
|||||
53 CATGGAGAAATTCAGTGCAGATTCTCTCTGCGCCCTCTCTCT 102
|||||
48 YrThrLeuAlaArgAspThrThrValLysProGlyAlaLysAspThr 64
|||||
103 ACACCTCTGCGCAGAGATACACAGCTCAAACTGGAGCCAAAAGACACA 152
|||||
65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
|||||
153 AAGGACTCTGCACCAAACTGCCAGACCTCTCCAGAGTTGGGGTGA 202
|||||
81 pGlnLeuIleThrPheGlnThrTrpGluAlaLeuTyrlsSerLysT 98
|||||

```

203 CCAACTCATCTGGACTGACATATGAGAAGCTCTATATAATCAAGA 252
 98 hSerSlnlySProLeuMetIleIleHisIleuAspGluCysProHis 114
 253 CAAGCAACAACCTTGATGATTATCATCATCTGGATGATGCCACAC 302
 115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
 303 AGTCAAGCTTTAAGAAAGTGTTCCTGAAATAATTAAGAAATCCAGAAAT 352
 131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
 353 GCGAGAGCAAGTTGTCTCCTCAATCGTTTATGAACAACAGACACAAC 402
 148 IseuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
 403 ACCTTCTCTCTGATGGCAGATATGCCCAAGATATGTTGTGGACCCA 452
 165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
 453 TCTCTACAGATTAGAGCCGATATCACTGGAGATATTCAAAACCGTCTCTA 502
 181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
 503 TGGTTACGAACCTGCAGATACAGCTCTGTGTTGCTTGACAACATGAAGAAG 552
 198 lAlaLeuLysLeuLeuLysThrGlu 205
 553 CTCTCAAGTTGCTGAAGACTGAC 575

seq_name: gb_est2:BF733142

seq_documentation_block:

LOCUS BF733142 755 bp mRNA linear EST 09-JAN-2001
 DEFINITION EST039 Human hepatocellular carcinoma subtracted cDNA library Homo
 sapiens cDNA clone P39 5', mRNA sequence.

ACCESSION BF733142
 VERSION BF733142.1 GI:12058378

KEYWORDS EST.
 SOURCE human.

ORGANISM human.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 755)
 Zhu, W., Duan, F., Liu, D., Ma, J., Bai, J. and Gao, F.

AUTHORS
 TITLE Suppression subtracted hybridization to identify differentially
 expressed genes of hepatocellular carcinoma and expressed sequence
 tags sequencing

JOURNAL
 COMMENT unpublished (2001)

CONTACT: Wuling Zhu
 Department of Nucleic Acid Research
 Institute of Digestive Disease

2 Jinda Road, Zhengzhou, 450003, Henan Province, PR.China
 Tel: 86 0371 3921444
 Fax: 86 0371 6960571

Email: wuling_z@hotmail.com

Human hepatocellular carcinoma cDNA research supported by Institute
 of Digestive Disease, Henan Medical University; cDNA insert
 sequencing: Genetech Biotechnology Company Limited. cDNA library
 construction: Department of Nucleic Acid Research, Institute of
 Digestive Disease.

Seq primer: 77.

FEATURES
 Source Location/Qualifiers

1..755
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="P39"
 /clone_lib="Human hepatocellular carcinoma subtracted cDNA
 library"
 /note="Organ: Liver"

BASE COUNT 237 a 171 c 158 g 189 t

alignment_scores:

Quality: 980.00 Length: 192
 Ratio: 5.131 Gaps: 0
 Percent Similarity: 99.479 Percent Identity: 99.479

alignment_block:

US-09-674-266A-181 x BF733142 ..

Align seg 1/1 to: BF733142 from: 1 to: 755

15 ProSerArgArgLeuThrGlnGlyArgTyrPValArgLysSerArgValAl 31
 3 CTTAGCCCGCCGACTCACACAGCCAGCGGTGAGAAATCCAGAGTTGC 52
 31 aMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerT 48
 53 CATGGAAGAAATTCAGAGTGCACGATTCCTTGTCTTGTGGCCCTCCT 102
 48 yThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
 103 ACACCTGGCCAGAGATACACACAGTCAAACTGAGCCAAAAGAGACACA 152
 65 LysAspSerArgProLysLeuProGlnThrLeuSerArgLysArgLys 81
 153 AAGGACTCTGCACCCAACTGCCCCAGACCCCTCTCCAGAGCTTGGGCTCA 202
 81 pGlnLeuIleTrrPThrGlnThrTyrGluGluAlaLeuTyrLysSerLysT 98
 203 CCAACTCATCTGGACTGACATATGAGAAGCTCTATATAATCAAGA 252
 98 hSerSlnlySProLeuMetIleIleHisIleuAspGluCysProHis 114
 253 CAAGCAACAACCTTGATGATTATCATCATCTGGATGATGCCACAC 302
 115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
 303 AGTCAAGCTTTAAGAAAGTGTTCCTGAAATAATTAAGAAATCCAGAAAT 352
 131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
 353 GCGAGAGCAAGTTGTCTCCTCAATCGTTTATGAACAACATGAAGAAG 402
 148 IseuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
 403 ACCTTCTCTCTGATGGCAGATATGCCCAAGATATGTTGTGGACCCA 452
 165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
 453 TCTCTACAGATTAGAGCCGATATCACTGGAGATATTCAAAACCGTCTCTA 502
 181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
 503 TGGTTACGAACCTGCAGATACAGCTCTGTGTTGCTTGACAACATGAAGAAG 552
 198 lAlaLeuLysLeuLeuLysThrGluLeu 206
 553 CTCTCAAGTTGCTGAAGACTGAAATTG 578

seq_name: gb_est2:BG501149

seq_documentation_block:

LOCUS BG501149 740 bp mRNA linear EST 27-MAR-2001
 DEFINITION 602346389P1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4668575 5',
 mRNA sequence.

ACCESSION BG501149
 VERSION BG501149.1 GI:13462666

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 740)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph. D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: DCTD/DTF

CDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LNCMI479 row: a column: 24
 High quality sequence stop: 738.

FEATURES

source

1. 740
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NIH_MGC_60"
 /tissue_type="adenoecarcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
 Site_1: Sfil (ggcgccctcgcc); Site_2: Sfil (ggccatagggc);
 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGCGCATTTGGCC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCGCGCGCGCATG-3' (30)BP-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies for
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
 library."

BASE COUNT 233 a 171 c 157 g 179 t

ORIGIN

alignment_scores:

Quality: 979.00 Length: 195
 Ratio: 5.021 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.487

alignment_block:

US-09-674-266A-181 x BG501149 ..

Align seg 1/1 to: BG501149 from: 1 to: 740

```

12 GYPRHISPROSERARGLEUTHRGINGLYAGTTPVALARGLYSE 28
|||||
14 GGACCGCATCTAGCCGCCGACTCACAGAGCGAGGTGGAGGAAATC 63
|||||
28 FARGVALAMETGLULYSILEPROVALSERALAPHEULEULEUVALA 45
|||||
64 CAGAGTTGCCATGGAGAAATTCAGGTGCACGATCTTGCCTCTTGCG 113
|||||
45 IALEUSERTYTHLEUALAARGASPTHRTVALYSPROGLVALALYS 61
|||||
114 CCCTCTCCACACTCTGGCGAGATACACAGCAACCTGAGCCAAA 163
|||||
62 LYSASPTHLYSASPSERARGPROLYSEUPROGLINTHLEUSERARG 78
|||||
164 AAGGACAAAGAGACTCTGACCCAAAGTCCGACGCCCTCTCCAGAG 213
|||||
78 YTPGLYASPGINLEULEITPTHTGINTHTYRGUGLUALAULTYRL 95
|||||
214 TTGGGCTGACCACTCATCTGAGCTCAGACATATAGAAGCTCATATA 263
|||||
95 YSSELYTHSERASNLYSPROLEUMELILEHISHISLEUASPGIU 111
|||||
264 ATCCAGAGCAAGCAAAACCTTGATGATATATCATCATCTGGATGAG 313
|||||
112 CYSPTHISSEIRGLALALEULYSLYSVALPHEALAGLUASNLYSGLU 128

```

```

|||||
314 TGCCACACAGTCAGAGCTTTAAGAAAGTGTTCGTAATAAGAAAT 363
|||||
128 EGLINLSLEUALAGLUGINPHEVALLEULEUASNLEUVALTYRGLUTH 145
|||||
364 CCAGAAATGGCAGAGCAGTT.GTCTCTCCATCATCTGGTTATGAACAA 412
|||||
145 HRASPLYSHISLEUSERPROASPGLYGINTYVALPROAATILEMCPHE 161
|||||
413 CTGACAAACACCTTTCTCTGATGCGCCAGTATGTCGCCAGCATATATG 462
|||||
162 VALASPPROSERLEUTHVALARGALASPILETHRGARGTYRSERAS 178
|||||
463 GTGACCCATCTGACAGTTAGACCGCATATTCACGTGAAATATTCAA 512
|||||
178 NARGLEUTYRALTYRGLUPROALAASPTHRALALEULEUASPSNM 195
|||||
513 TCGTCTCATCTGATAGCAAGCTGCAGATACAGCTCTGTGTTGACAA 562
|||||
195 ELYSLYALALEULYSLEULEULYSTHRTGLUEU 206
|||||
563 TGAAGAAAGCTCTCAAGTTCCTGAAGACTGAATTG 597
|||||

```

seq_name: gb_est2:BI770077

seq_documentation_block:

LOCUS BI770077 788 bp mRNA linear EST 25-SEP-2001
 DEFINITION 603053412F1 NIH_MGC_122 Homo sapiens CDNA clone IMAGE:5202866 5',
 mRNA sequence.

ACCESSION BI770077

VERSION BI770077.1 GI:15761655

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 788)
 NIH-MGC <http://mgc.ncl.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph. D.

COMMENT

Email: cgabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LNCMI508 row: h column: 03
 High quality sequence stop: 786.

FEATURES

source

1. 788
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5202866"
 /clone_lib="NIH_MGC_122"
 /lab_host="DH10B"
 /note="Organ: pooled lung and spleen; Vector: PCMV-SF0R6;
 Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
 anonymous pool of 24 week female lung, 16 week female
 spleen, and 20-22 week male spleens. Library is oligo-dT
 primed and directionally cloned (EcoRV site is destroyed
 upon cloning). Average insert size 1.4 kb, insert size
 range 1-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 026. Note:
 this is a NIH_MGC library."

BASE COUNT 248 a 178 c 165 g 197 t

ORIGIN

alignment_scores:

Quality: 976.00 Length: 194
Ratio: 5.083 Gaps: 1
Percent Similarity: 98.969 Percent Identity: 98.969

alignment_block:

US-09-674-266a-181 x B1770077 ..

Align seg 1/1 to: B1770077 from: 1 to: 788

```

14 HistproserArgArgLeuThrGlnGlyArgTyrPylAlaGlySerArgVa 30
3 CATCTAGCCCGCCGACTCACACAAGCAGGTGGGTGAGGAATCCAGAGT 52
30 LAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaLeu 46
53 TGCCATGAGAGAAAATTCACAGTGCAGCATCTTGCTCTGTGGCCCTC 102
47 SerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 63
103 TCCTACACTGTGGCCAGAGATACCACAGTCAACCTGGAGCCAAAAGGA 152
63 PThrLysAspSerArgProLysLeuProGlnThrLeuSerArgLysTyr 80
153 CACAAGAGACTCTGACCCCAACTGCCAGACCCCTCTCCAGAGGTGG 202
80 LysAspGlnLeuIleTyrThrGlnThrTyrGluGluAlaLeuTyrLys 96
203 GTGACCAACTATCTGACACTCAGACATATGAGAAGCTCTATATTAAT 252
97 LysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluCys 113
253 AAGACAAGCAACAACCTTGATGATTATTCATCACCCTGGATGAGTGC 302
113 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGln 130
303 ACACAGTCAAGCTTAAAGAAAGTTTGTGAAATTAAGAAATCCAGA 352
130 LysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrAsp 146
353 AATTGGCAGACAGTCTTCTCCTCCTCAATCTGTGTTATGAACAAC 402
147 LysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 163
403 AAACACCTTCTCCTGATGGCCAGATGTCCCAAGATTATGTTGTGA 452
163 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArg 180
453 CCCATCTCTGACAGTACAGCCGATATCATCAGGAAGATATTCAAAT 502
180 eUtyrAlaTyrGluProAlaAspThrAlaLeuLeuAspAsnMetLys 196
503 TCATAGCTTACGAACCTGCAGATACAGCTGTGGCTTGACACACATG 552
197 LysAlaLeuLysLeuLeuLysThrGluLeu 206
553 AAAGCTCTCAAGTGTCTGACAGACTGAATTG 582

```

seq_name: gb_est1:AW956284

seq_documentation_block:

LOCUS AW956284 721 bp mRNA linear EST 01-JUN-2000
DEFINITION EST368354 MAGE resequencences, MAGD Homo sapiens cDNA, mRNA sequence.

ACCESSION AW956284
VERSION AW956284.1 GI:8145967

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 721)
Hegde,P., Qi,R., Abernathy,R., Dharap,S., Gaspard,R., Gay,C., Holt

,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.

TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray

JOURNAL Unpublished (2000)

CONTACT: John Quackenbush

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528
Fax: 301 838 0208

Email: johnq@igrr.org

Plate: 90

Seq primer: Reverse.

Location/Qualifiers

source 1..721

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="MAGE resequencences, MAGD"

/note="Vector: pBluescriptSKm"

BASE COUNT 224 a 173 c 148 g 175 t 1 others

ORIGIN

alignment_scores:
Quality: 975.00 Length: 194
Ratio: 5.078 Gaps: 0
Percent Similarity: 98.969 Percent Identity: 97.938

alignment_block:

US-09-674-266a-181 x AW956284 ..

Align seg 1/1 to: AW956284 from: 1 to: 721

```

13 ProHisProSerArgArgLeuThrGlnGlyArgTyrPylAlaGlySerArg 29
6 CGGACTCTGCCGAGACGACTCACACAAGCAGGTGGGTGAGGAATCCAG 55
29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaL 46
56 AGTTGCCATGAGAAATTCACAGTGTGACAGTCTTGCTCTGTGGCC 105
46 eUSeTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 62
106 TCTCTACACTCTGGCCAGAGATACCACAGTCAACCTGGAGCCAAA 155
63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 79
156 GACACAAAGGACTCTGCACCAAACTGCCCAAGCCCTCCAGAGGTG 205
79 pGlyAspGlnLeuIleTyrThrGlnThrTyrGluGluAlaLeuTyr 96
206 GGGTACCAACTCATCTGAGACTCAGACATATGAAGAACTCTATTAAT 255
256 eUtyrThrSerAsnLysProLeuMetIleIleHisLysLeuAspGlu 112
113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 129
306 CCACACAGTCAAGCTTAAAGAAAGTGTTCCTGAAATTAAGAAATCA 355
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 146
356 GAAATTTGGCAGACAGAGTTGTCTCTCTCAATCTGTTATGAACAAC 405
146 sPlyHisLysLeuSerProAspGlyGlnTyrValProArgIleMetP 162
406 ACAAAACACTTCTCTGATGGCCAGTATGTCCAGAGATTATGTTGTT 455
163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 179
456 GACCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 505
179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuAspAsnMet 196

```

506 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACACATGA 555
 196 YtYsAlaLeuLysLeuLeuLysThrGluLeu 206
 556 AGAAGCTCTCAAGTTGCTGAGAGCTGAATTG 587

seq_name: qb_est2:BG285246

seq_documentation_block:

LOCUS BG285246 866 bp mRNA linear EST 21-FEB-2001
 DEFINITION 602409569F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4539353 5',
 mRNA sequence.

ACCESSION BG285246
 VERSION BG285246.1 GI:13037011

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov

Tissue Procurement: DCTD/DRP
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLML10467 row: 1 column: 18
 High quality sequence stop: 750.

FEATURES Location/Qualifiers

source

1..866
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4539353"
 /clone_1lb="NIH_MGC_91"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
 Average insert size 1.4 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 270 a 194 c 189 g 212 t 1 others
 ORIGIN

alignment_scores:
 Quality: 971.50 Length: 194
 Ratio: 5.086 Gaps: 1
 Percent Similarity: 98.454 Percent Identity: 98.454

alignment_block:
 US-09-674-266a-181 x BG285246 ..

Align seg 1/1 to: BG285246 from: 1 to: 866

14 HsPpSer...ArgArGLeuThrGngLyArgTrpValAlaGlySerAr 29
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 18 CAGCGCTCGCGCGCGCGACTCACACAGGAGGTGGTGAAGAAATCCAG 67
 29 gValAlaMeGluLysIleProValSerAlaPheLeuLeuValAlaL 46
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 68 AGTGGCAGAGAGAAATTCACAGTTCAGCATTCCTGCTCTTGCGCC 117
 46 euseTyTrhLeuAlaArgAspThrThrValLysProGlyAlaLys 62
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 118 TCTCTACACTCTGCGCAGAGATACCAAGTCAAACTGGAGCCAAAAG 167

63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 168 GACACAAAGAGACTCTGCAGCCCAACTGCCAGACCTCTCCAGAGTTG 217
 79 pGlyAspGlnLeuIleTrpThrGlnThrTyrgLugLualLeuTyLys 96
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 218 GGGTGACCAACTCATCTGCAGCTCAGACATATGAAGAAGCTCTATATAAT 267
 96 eTrpThrSerAsnLysProLeuMetIleIleHisThrLeuAspGluG 112
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 268 CCAAGACAAAGCAACAAACCTTATATATATATATATATATATATAT 317
 113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 129
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 318 CCACACAGTCAAGCTTTAAAGAAAGTTTGCCTGAAGAAATTAACAAATCC 367
 129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgLutThra 146
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 368 GAAATGGCAGACAGCTTGTCTCCTCAATCTGTATGAACAACATG 417
 146 sPrpHisLeuSerProAspGlyGlnTrpValProArgIleMetPheVal 162
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 418 ACAACACCTTCTCTCTGATGCGCAGATGTCCCGAGGATATGTTTGT 467
 163 AspProSerLeuThrValAlaArgAlaAspIleThrGlyArgTySerAsn 179
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 468 GACCACTCTCTGACAGTTAGAGCCGATATCAGCGAAGATATCAATG 517
 179 gLeuTyAlaTyrgLupProAlaAspThrAlaLeuLeuAspAsnMet 196
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 518 TCTCTATGCTTACGAACCTGCAGATACAGCTGTGCTTGACACATGA 567
 196 YtYsAlaLeuLysLeuLeuLysThrGluLeu 206
 568 AGAAGCTCTCAAGTTCAGTGAAGACTGAATTG 599

seq_name: qb_est2:BM129777

seq_documentation_block:

LOCUS BM129777 581 bp mRNA linear EST 27-NOV-2001
 DEFINITION 1f23a08.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 cDNA 5' similar to TR:095994 095994 SECRETED CEMENT GLAND PROTEIN
 XAG-2 HOMOLOG. ; mRNA sequence.

ACCESSION BM129777
 VERSION BM129777.1 GI:17124329

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
 Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
 Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
 Schmitt,A., Theising,B., Rittler,E., Ronko,T., Bennett,J., Cardenas
 M., Gibbons,M., McCann,R., Cole,R., Tsagarelshvili,R., Williams,T.,
 Jackson,Y. and Bowers,Y.
 Endocrine Pancreas Consortium
 Other-ESTs: 1f23a08.x1

CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@iobp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center for information on
 obtaining a clone please contact: Juliana Brown
 (brownjefas.harvard.edu)
 High quality sequence stop: 429.

FEATURES
source

Location/Qualifiers
1..581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Mellon Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation: average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 185 a 140 c 118 g 138 t
ORIGIN

alignment_scores: Quality: 971.00 Length: 189
Ratio: 5.138 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266a-181 x BM129777 ..

Align seg 1/1 to: BM129777 from: 1 to: 581

```

18 ATGLeuThGlnGlyValGTPValArgLysSerArgValAlaMetGluLys 34
1 CGACTCACACAAAGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGAA 50
34 sileProValSerAlaPheLeuLeuValAlaLeuSerTyrThrLeuA 51
51 AATTCAGAGTGCAGCATCTCTGCTCTGTGGCCCTCTCTACACTCTGG 100
51 laArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSer 67
101 CCAGAGATACCAACAGTCAAACTGGAGCCAAAAGACACAAAGACTCT 150
68 ArgProLysLeuProGlnThrLeuSerArgGlyTyrGlyAspGlnLeuL 84
151 CGACCCAAACTGCCAGACCTCTCCAGAGTTGGGGTGACCAACTCAT 200
84 eTPThrGlnThrTyrGlnGluAlaLeuTyrLysSerLysThrSerAsnL 101
201 CTGAGATCGACATATGAGAAAGCTCTATTAATTCACAGACAAAGCACA 250
101 ySPProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAla 117
251 AACCCCTTGATGATTATTCATCACTGATGAGTGCACACACTCAAGCT 300
118 LeuLysLysValAlaPheAlaGluAsnLysGluIleGlnLysLeuAlaGlu 134
301 TTAAACAAAGGTGGTCTGAATAAAGAAATCCAGAAATTTGGCAGAGCA 350
134 nPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerP 151
351 GTTTGGCTCTCTCAATCTGGTTTATGAAACACTGACAAACACTTTCTC 400
151 roAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuThr 167
401 CTGATGGCCAGTATGTCCCGAGAGATATATTGTGTGACCATCTCTGACA 450

```

```

168 ValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGlu 184
451 GTTAGAGCCGATATCACTGGAAAGATATTCAAATCGTCTCATCTTACGCA 500
184 uProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysL 201
501 ACCTGCAGATACAGCTCTGTGCTGTGACAAACATGAGAAAGACTCAAGT 550
201 euleuLysThrGluLeu 206
551 TGCTGAAGACTGAATTG 567
seq_name: gb_est2:BE796856

```

seq_documentation_block:
LOCUS BE796856 948 bp mRNA linear EST 20-SEP-2000
DEFINITION 601588166F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3942112 5',
mRNA sequence.
ACCESSION BE796856
VERSION BE796856.1 GI:10218156
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: DCTD/DMP

cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LCM795 row: d column: 17
High quality sequence stop: 748.

FEATURES
source

Location/Qualifiers
1..948
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3942112"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 296 a 225 c 207 g 220 t
ORIGIN

alignment_scores: Quality: 962.00 Length: 187
Ratio: 5.144 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266a-181 x BE796856 ..

Align seg 1/1 to: BE796856 from: 1 to: 948

```

20 ThGlnGlyArgTTPValArgLysSerArgValAlaMetGluLysIlePr 36
|||||
3 ACACAAAGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGAGAAATATTC 52

```


178 nargLeuYrAlaTyrgluProAlaaspThrAlaLeuLeuAspAsnM 195
|||||
496 TCGCTCTATGCTTACGAACCTGCAGATACAGCTGTTGCTTGACACACA 545
195 etLysLysAlaLeuLysLeuLysThrgluLeu 206
|||||
546 TGAAGAAAGCTCTCAAGTTGCTGAGACCTCAATTG 580
seq_name: gb_est2:BE870718
seq_documentation_block:
LOCUS BE870718 734 bp mRNA linear EST 20-OCT-2000
DEFINITION 60148570F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852448 5',
mRNA sequence.
ACCESSION BE870718
VERSION BE870718.1 GI:10319507
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 734)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9574 row: 1 column: 17
High quality sequence stop: 723.
Location/Qualifiers
source 1..734
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3852448"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: Noti;
Site_2: SalI; Cloned unidirectional. Primer: Oligo dT.
Average insert size 1.8 Kb. Library constructed by Life
Technologies."
BASE COUNT 231 a 163 c 156 g 184 t
ORIGIN
alignment_scores:
Quality: 957.00 Length: 186
Ratio: 5.145 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-266a-181 x BE870718 ..
Align seg 1/1 to: BE870718 from: 1 to: 734
21 GlnGjArGrTrpValArgLysSerArgValAlaMetGluLysIleProVa 37
|||||
2 CAAGGCAAGGTGGGTGAGGAATCCAGAGTTGCATGGAGAAATTCAGT 51
|||||
37 1SerAlaPheLeuLeuLeuValAlaLeuSerTyTrhLeuAlaArqAsp 54
|||||
52 GTAGACATTTTGTCTTGTGGCCCTCTCTACACTGTGCCAGACATA 101
|||||
54 hrThrValLysProGlyAlaLysLysAspThrLysAspSerArgProLys 70
|||||

102 CCACAGTCAAAACCTGGAGCCAAAAAGACACAAAGACTCTCGACCCAAA 151
71 LeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuLleTrpThrG 87
|||||
152 CTGCCCCAGACCTCTCCAGAGGTTGGGGTGACCACACTCATCTGACTCA 201
87 nrThrYrGluGluAlaLeuTyrlLysSerLysTrhSerAsnLysProLeuM 104
|||||
202 GACATATGAAGAAGCTCTATATAATCCAAAGACACAGCAACAAACCTTGA 251
104 etLleLleHsHsLeuAspGluCysProHisSerGlnAlaLeuLysLys 120
|||||
252 TGAATTATTCATCACTTGAGTAGTGGCCACACAGTCAAGCTTTAAAGAAA 301
121 ValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGlnPheVal 137
|||||
302 GTGTTTGCTGAAAATTAAGAAATCCAGAAATTGGCAGAGCATTTGTCT 351
137 uLeuAsnLeuValTyrgluThrAspLysHsLeuSerProAspLys 154
|||||
352 CCGCAATCTGTGTTTANGAAACAACTGACAAACACTTCTCTGATGGCC 401
154 InTyValProArgLleMetPheValAspProSerLeuThrValArgAla 170
|||||
402 AGTAGTCCCGCAGAGATTATGTGTGGACCCACATCTGACAGTTAGAGCC 451
171 AspLleThrGlyArgTySerAsnArgLeuTyrlAlaTyrgluProAlaAs 187
|||||
452 GATATCACTGGAAGATATTCAAAACCGCTCTATGCTTACCAACCTGAGA 501
187 pThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysLeuLysTr 204
|||||
502 TACAGCTCTTGTCTTACAAACATGAAGAAGCTCTAAGTTCTGTGAGA 551
204 hrGluLeu 206
|||||
552 CTGAATTG 559
seq_name: gb_est2:BM006356
seq_documentation_block:
LOCUS BM006356 758 bp mRNA linear EST 30-OCT-2001
DEFINITION 603615055F1 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5421109 5',
mRNA sequence.
ACCESSION BM006356
VERSION BM006356.1 GI:16520710
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 758)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CK1876 row: m column: 14
High quality sequence stop: 754.
Location/Qualifiers
source 1..758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5421109"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site.1: XhoI;
Site.2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the Laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 238 a 176 c 156 g 188 t
ORIGIN

alignment_scores: Quality: 948.00 Length: 193
 Ratio: 5.070 Gaps: 1
Percent Similarity: 96.891 Percent Identity: 96.891

alignment_block:
US-09-674-266A-181 x BM006356 ..

Align seg 1/1 to: BM006356 from: 1 to: 758

```
14 HISPProSerAArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
   |||
4 CATCTAGCCCGCCGACTCACACAAGC.....AGAGT 35
30 LAlMeGluLysIleProValSerAlaPheLeuLeuValAlaLeuS 47
   |||
36 TGCCATGGAGAAATCCAGTGCAGCATCTCTCTGCGCCCTCT 85
47 eRTYThrLeuAlaArgSPThrThrValLysProGlyAlaLysLysASP 63
   |||
86 CCTACACTCTGGCCAGAGATACCAAGTCAACCTGGAGCCAAAAGGAC 135
64 ThrLysASPserArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
   |||
136 ACAAGGACTCTCGACCCAACTGCCGAGACCTCTCCAGAGTTGGGG 185
80 YASPGLneutIleTrrHrgInThrGlnGluAlaLeuTrrLysSerL 97
   |||
186 TGACCAACTACTGTGACTCAGACATATGAAAGCTCTATATTAATCCA 235
97 YSTHrSerSPnLysProLeuMetIleIleHISLysLeuASPGLucysPro 113
   |||
236 AGACAGACAAACCCCTTGATGATTAATTCACCTTGATGAGTCCCA 285
114 HISserGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
   |||
286 CACAGTCAAGCTTAAAGAAAGTGTTCGTAATAAATAAGAAATCCAGAA 335
130 sLeuAlaGlnIlnPheValLeuLeuAsnLeuValTyrGlnThrHraSP 147
   |||
336 ATTTGGCAGAGAGTTTGTCTCTCAATCGTTGATGAAACAACACTGACA 385
147 YSHLSLeuSerProASPGLysInTyrValProArgIleMetPheValASP 163
   |||
386 AACACCTTTCCTCGATGCGCAGTATGCCAGAGATTATGTTTGTGAC 435
164 ProSerLeuThrValArgAlaASPleThrGlyArgTyrSerAsnArgLe 180
   |||
436 CCATCTCTGACAGTTAGACCGCATATCACTGAGATATTCAAAACCTGCT 485
180 uTYrAlaTyrGlnProAlaASPThrAlaLeuLeuAspAsnMetLysL 197
   |||
486 CTATGCTTACGAACTGCAGATACAGCTGTGCTGCTGACACATGAAGA 535
197 YSAlaLeuLysLeuLeuLysThrGlnLeu 206
   |||
536 AAGCTCTCAAGTGTGCTGAAGACTGAATTG 564
```

OM of: US-09-674-266a-181 to: GenEmbl:* out_format: pfs
Date: Oct 8, 2002 11:42 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+2n.spool -DEV=x1h
-O=/gen2.1/USPTO.spool/US09674266/Runat_08102002.091409_27005/app-query.fasta.1.271
-DB=GenEmbl -OPT=fastap -SUFFIX=olip2n.rge -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELDP=6.000 -DELEXT=7.000 -START=1 -MAREXT=0.150
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs -NOM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09674266_EGNT_1.3973 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-674-266a-181
Query length: 206
Database: GenEmbl:*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 1915.830000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Score List:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
gb_pat:AX011612	+	206.00	3944.17	2.5e-211	1020	AX011612 Sequence 8 from Patent
gb_pat:AX015056	+	206.00	3944.17	2.5e-211	1020	AX015056 Sequence 265 from Patent
gb_pat:AF088867	+	195.00	3712.51	1.5e-199	956	AF088867 Homo sapiens putative
gb_pat:AF038451	+	194.00	3709.14	1.5e-198	1077	AF038451 Homo sapiens secreted
gb_pat:AF015926	+	194.00	3709.14	1.5e-198	1077	AF015926 Homo sapiens XAC-2 bc
gb_pat:BC015503	+	193.00	3695.10	1.9e-197	836	BC015503 Homo sapiens, anterior
gb_pat:AR123828	+	192.00	3675.45	2.3e-196	875	AR123828 Sequence 1 from patent
gb_pat:AX341084	+	184.00	3524.65	5.8e-188	557	AX341084 Sequence 1331 from Patent
gb_pat:AX231601	+	178.00	3409.20	1.6e-181	543	AX231601 Sequence 7 from Patent
gb_pat:AF007291	+	178.00	3406.00	2.4e-181	838	AF007291 Homo sapiens secreted
gb_pat:AR177367	+	178.00	3400.83	4.6e-181	1689	AR177367 Sequence 61 from Patent
gb_pat:AX341234	+	177.00	3389.15	2.1e-180	603	AX341234 Sequence 1481 from Patent
gb_pat:AX341953	+	161.00	3081.82	2.7e-163	525	AX341953 Sequence 2200 from Patent
gb_pat:AX351468	+	160.00	3060.44	4.2e-162	699	AX351468 Sequence 215 from Patent
gb_pat:AX067336	+	134.00	2558.57	3.8e-134	778	AX067336 Sequence 40 from Patent
gb_pat:AX192616	+	133.00	2544.04	2.4e-133	409	AX192616 Sequence 183 from Patent
gb_pat:AX341884	+	121.00	2310.67	2.4e-120	544	AX341884 Sequence 2131 from Patent
gb_pat:AX341524	+	120.00	2291.40	2.9e-119	544	AX341524 Sequence 1771 from Patent
gb_pat:AX341388	+	120.00	2290.76	3.1e-119	593	AX341388 Sequence 1635 from Patent
gb_pat:AX340668	+	115.00	2195.48	6.3e-114	512	AX340668 Sequence 915 from Patent
gb_pat:AX341186	+	108.00	2059.62	2.3e-106	583	AX341186 Sequence 1433 from Patent
gb_pat:AR123857	+	107.00	2024.35	1.9e-105	386	AR123857 Sequence 35 from Patent
gb_pat:AX340984	+	106.00	2041.49	3.1e-104	554	AX340984 Sequence 1231 from Patent
gb_pat:AX302898	+	103.00	1967.66	3.1e-101	321	AX302898 Sequence 218 from Patent
gb_pat:AX156194	+	103.00	1963.98	5.0e-101	529	AX156194 Sequence 51 from Patent
gb_pat:AR144213	+	102.00	1941.72	8.6e-100	793	AR144213 Sequence 78 from Patent
gb_pat:AR176400	+	102.00	1941.72	8.6e-100	793	AR176400 Sequence 1386 from Patent
gb_pat:AX341139	+	101.00	1925.86	6.6e-99	499	AX341139 Sequence 1546 from Patent
gb_pat:AX341299	+	93.00	1773.04	2.1e-90	415	AX341299 Sequence 2 from Patent
gb_pat:AX322727	+	80.00	1524.37	1.5e-76	322	AX322727 Sequence 1476 from Patent
gb_pat:AX341229	+	76.00	1443.70	4.7e-72	524	AX341229 Sequence 32 from Patent
gb_pat:AR166839	+	66.00	1252.95	2.0e-61	401	AR166839 Sequence 32 from Patent
gb_pat:AX192465	+	66.00	1252.95	2.0e-61	401	AX192465 Sequence 149 from Patent
gb_pat:AR177410	+	62.00	1169.58	8.8e-57	301	AR177410 Sequence 1363 from Patent
gb_pat:AX192553	+	58.00	1100.88	5.9e-53	301	AX192553 Sequence 120 from Patent
gb_pat:AX341116	+	57.00	1076.67	1.3e-51	588	AX341116 Sequence 163 from Patent
gb_pat:AX261879	+	56.00	1065.32	5.6e-51	201	AX261879 Sequence 1530 from Patent
gb_pat:AX341514	+	55.00	1042.00	1.1e-49	348	AX341514 Sequence 1761 from Patent

gb_pat:AR123858	+	53.00	1002.94	1.7e-47	373	AR123858 Sequence 36 from Patent
gb_pat:AX106428	+	52.00	989.95	8.9e-47	159	AX106428 Sequence 209 from Patent
gb_pat:AX140719	+	52.00	989.95	8.9e-47	159	AX140719 Sequence 209 from Patent
gb_pat:AX200579	+	52.00	989.95	8.9e-47	159	AX200579 Sequence 209 from Patent
gb_pat:AX267235	+	52.00	989.95	8.9e-47	159	AX267235 Sequence 209 from Patent
gb_pat:AR123859	+	51.00	970.73	1.0e-45	158	AR123859 Sequence 37 from Patent
gb_pat:AX062783	-	50.00	948.53	1.8e-44	235	AX062783 Sequence 410 from Patent

seq_name: gb_pat:AX011612

seq_documentation_block:

LOCUS AX011612 1020 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 8 from Patent WO9555858.

ACCESSION AX011612

VERSION AX011612.1 GI:9998136

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

BASE COUNT

ORIGIN

alignment_scores:

Quality: 206.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x AX011612

Align seg 1/1 to: AX011612 from: 1 to: 1020

1	ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerar	17	
57	AGACTGACGTGCTGCTGCACACTGACGCGCTGACCGATCTACCG	106	
17	gATgLeuThGInGInGlyrTrpValArgGlySerArgValAlaMetGlu	34	
107	CCGACTCACAACAAGAGGCGGAGAAATCCAGAGTTCACATGAGAGA	156	
34	ysIlePProValSerAlaPheLeuLeuValAlaLeuSerTyThrLeu	50	
157	AAATTCAGTCTGACGATTCCTGCTTGTGGCCCTCTCCACACTCTG	206	
51	AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe	67	
207	GCCAGAGATACACAGCAAACTGAGCCAAAAGAGACAAAGAGACTC	256	
67	rATgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeu	84	
257	TCGACCCAACTGCCAGACCTCTCCAGAGTGGGGGACCAACTCA	306	
84	leTrpThrGInThrTyGInGluAlaLeuTyTrpLysSerLysThrSerAsn	100	
307	TCGACATCAGACATATGAGAGACTTATATATATCCAGACAGCAAGAC	356	
101	LysProLeuMetIleIleHisLysLeuAspGluCysProHisSerGlnAl	117	
357	AAACCTTGATGATTATTCATCACTGTGATGAGTGCACACAGTCAAGC	406	

```

117 aleuylslysva1phea1agluasnlysgluileglnlysleuallaglug 134
|||||
407 TTTAAAGAAAGTGTTCGTAAGAAATCCAGAAATGGCAGAGC 456
|||||
134 lnphvalleuasnleuvaltyrgluthrthrasplyshsleuser 150
457 AGTTTGTCCTCCCAATCTGTTTATGAACAACACGTAACACCTTTCT 506
|||||
151 Proaspg1glnTyValProArq1IemEtpheValasProserleuth 167
|||||
507 CCTGATGGCAGATATGTCCTCCAGATATGTTGTTGACCATCTCTGAC 556
|||||
167 rValArGa1aasp1IeThrGlyArqTySerAsnArqLeuTyra1aTyrg 184
|||||
557 AGTTAGACCGCATATCAGTAAGAAATATCAAAACGCTCTATGCTTACG 606
|||||
184 lnPro1a1sPthRAleuLeuLeuAspAsnMetLysLysAlaLeuLys 200
607 AACCTGCAGATACAGCTCTGTTGTTGACACATGAAGAAAGCTCTCAAG 656
201 LeuLeuLysThrgLudeu 206
|||||
657 TTGCTGAACACTGAATTG 674

seq_name: gb_pat:AX015056

seq_documentation_block:
LOCUS AX015056 1020 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 265 from Patent WO9953040.
ACCESSION AX015056
VERSION AX015056.1 GI:10041195
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1020)
Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Plarsky,C.
Human nucleic acid sequences from ovarian tumor tissue
JOURNML Patent: WO 9953040-A-265 21-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PIARSKY CHRISTIAN (DE)
FEATURES
source 1..1020
Location/Qualifiers
BASE COUNT 341 a 214 c 218 g 247 t
ORIGIN
alignment_scores:
Quality: 206.00 Length: 206
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266a-181 x AX015056 ..

Align seg 1/1 to: AX015056 from: 1 to: 1020
1 ArgLeuSerCyAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
|||||
57 AACACTACAGTGTGCTGCGACACTACAGCGCTCTGGACCCGATCTAGCCG 106
|||||
17 gArqLeuThrgInGlyArqTyPValArqLysSerArqValAlaMetGluL 34
|||||
107 CCGACTCACACAAGGAGGTGGGTGAGAAATCCAGATTCGCATCGAGACA 156
|||||
34 yAl1eProValSerAlaPheLeuLeuValAlaLeuSerTyThrLeu 50
|||||

```

```

157 AATTCAGCTGTCAGCATTCCTTGCTCTTGCGCCCTCTCTACACTCTG 206
|||||
51 AlaArqAspThrThrValLysProGlyAlaLysLysAspThrLysAsp 67
|||||
207 GCCAGAGATCCACAGTCAAACTGAGGCCAAAAGAGACACAAAGAGCTC 256
|||||
67 rArqProLysLeuProGlnThrLeuSerArqGlyTyrgLysAspGlnLeu 84
|||||
257 TCGACCCAAAGTCCGCCACAGACCTCTCCAGAGTTGGGTGACCAACTCA 306
84 lArqThrgInThrTyrgLugLualalaLeuTyrrLysSerLysThrSerA 100
|||||
307 TCTGACTCAGACATATGAAGAGCTATATTAATCCAAAGCAAGCAAC 356
|||||
101 LysProLeuMet1Ie1IeHisLysLeuAspGluCysProHisSerGlnAl 117
|||||
357 AAACCTTGATGATATATTCATCAGTTCGATGATGCGCCACACAGTCAAC 406
|||||
117 aleuylslysva1phea1agluasnlysgluileglnlysleuallaglug 134
|||||
407 TTTAAAGAAAGTGTTCGTAAGAAATCCAGAAATGGCAGAGC 456
|||||
134 lnphvalleuasnleuvaltyrgluthrthrasplyshsleuser 150
457 AGTTTGTCCTCCCAATCTGTTTATGAACAACACGTAACACCTTTCT 506
|||||
151 Proaspg1glnTyValProArq1IemEtpheValasProserleuth 167
|||||
507 CCTGATGGCAGATATGTCCTCCAGATATGTTGTTGACCATCTCTGAC 556
|||||
167 rValArGa1aasp1IeThrGlyArqTySerAsnArqLeuTyra1aTyrg 184
|||||
557 AGTTAGACCGCATATCAGTAAGAAATATCAAAACGCTCTATGCTTACG 606
|||||
184 lnPro1a1sPthRAleuLeuLeuAspAsnMetLysLysAlaLeuLys 200
607 AACCTGCAGATACAGCTCTGTTGTTGACACATGAAGAAAGCTCTCAAG 656
657 TTGCTGAACACTGAATTG 674

seq_name: gb_pr:AF088867

seq_documentation_block:
LOCUS AF088867 956 bp mRNA linear PRI 02-JAN-2000
DEFINITION Homo sapiens putative secreted protein XAG mRNA, complete cds.
ACCESSION AF088867
VERSION AF088867.1 GI:6652811
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 956)
Zhang,J.S. and Smith,D.I.
Human homolog of XAG is differentially expressed in tumors
JOURNML Unpublished
REFERENCE
2 (bases 1 to 956)
Zhang,J.S. and Smith,D.I.
Direct Submission
TITLE
Submitted (30-AUG-1998) pathology and Lab. Medicine, Mayo Clinic,
200 SW 1st St., Rochester, MN 55905, USA
FEATURES
source 1..956
Location/Qualifiers
1..956
/organism="Homo sapiens"
/db_xref="taxon:9606"
148..675
/note="differentially expressed in tumors; similar to
Xenopus cement gland-specific secreted protein"
/codon_start=1
/product="putative secreted protein XAG"
/protein_id="AAF22484.1"

```

/db.xref="GI:6652812"
 /translation="MERIPVSAFLILVALSTLARDTVKPKAKDTPKPKLPOTL
 SRGWDOLIMTQTYEALYKSKTNPLMIITHHIDEPHSQALKVFAENKEIOKLAE
 QEVLLNLYETTDKHLSPDQYVPRIMEVPSLVRADITGRYSNRLYAYEPADTALL
 LDNMKKALKLKTLEL"
 BASE COUNT 310 a 202 c 202 g 242 t
 ORIGIN

alignment_scores:
 Quality: 195.00 Length: 195
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-266a-181 x AF088867 ..

Align seg 1/1 to: AF088867 from: 1 to: 956

```

12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
|||||
88 GGACCCGATCTAGCCGCGACTCACACAGGCGAGGTGGGTGAGGAATC 137
28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAla 45
|||||
138 CAGAGTTGGCATTGAGAAATTCAGTGTGACATCTTCTCTCTGTG 187
45 lAluSerTrpThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
|||||
188 CCTCTCTTACACTGTGGCCAGATACCAAGTCAACCTGGAGCCAAA 237
62 LysAspThrLysAspSerArgProLysLeuProGlnThrLysSerArgL 78
|||||
238 AAGGACAAAGAGCTCTGACCCCAACAGTCCCAAGCCCTCTCCAGAG 287
78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTrpGluGluAlaLeuTrp 95
|||||
288 TTGGGGGTGACCAACTCATCTGAGCTCAGACATATGAAGAAGCTTATA 337
95 ySerSerThrSerAsnLysProLeuMetIleIleHisLysLeuAspGlu 111
|||||
338 AATCCAAAGCAAGCAACAAACCTTGATGATATTCATCATCTTGATGAG 387
112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGlu 128
|||||
388 TGCCCAACAGTCAAGCTTTAAAGAAAGTGTTCCTGAAATAAAGAAAT 437
128 eGlnLysLeuAlaGluLysPheValLeuLysAsnLeuValTrpGluThr 145
|||||
438 CCAGAAATGGCAGAGCAGTGTCTCTCCATCATCTGGTTATGAACA 487
145 hAspLysHisLysSerProAspGlyGlnTrpValProArgIleMetPhe 161
|||||
488 CTGACAAACACCTTCTCTGATGGCCAGTATGTCCTCCAGGATTAAT 537
162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTrpSer 178
|||||
538 GTTGACCCCTCTCTGACACTTAGAGCCGATATCAGTGAAGATATTCAA 587
178 nArgLeuTrpAlaTrpGluProAlaAspThrAlaLeuLeuAspAsnM 195
588 TCGTCTCTAATGCTTAGCAACCTGCAATACAGCTCTGTTGCTTGACA 637
195 eTrpLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
638 TGAAGAAAGCTCTCAAGTTGCTGAAGACGTAATTTG 672

```

seq_name: gb_pr:AF038451

seq_documentation_block:

LOCUS AF038451 1077 bp mRNA linear PRI 03-DEC-1998
 DEFINITION Homo sapiens secreted cement gland protein XAG-2 homolog (hAG-2/R)
 mRNA, complete cds.

ACCESSION AF038451
 VERSION AF038451.1 GI:3779225
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Thompson, D.A. and Weigel, R.J.
 TITLE 1 (bases 1 to 1077)
 JOURNAL hAG-2, the human homologue of the Xenopus laevis cement gland gene
 MEDLINE hAG-2, is coexpressed with estrogen receptor in breast cancer cell
 99009231 lines
 REFERENCE 2 (bases 1 to 1077)
 AUTHORS Thompson, D.A.
 TITLE Direct Submission
 JOURNAL Submitted (09-DEC-1997) Surgery, Stanford University, 1201 Welch
 Road, MSLS Building, Room P228, Stanford, CA 94305, USA
 FEATURES
 source
 1..1077
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="MCF7"
 /tissue_type="breast"
 /note="Isolated from estrogen receptor-positive breast
 adenocarcinoma cell line"
 1..1077
 /gene="hAG-2/R"
 59..586
 /gene="hAG-2/R"
 /note="similar to Homo sapiens protein hAG-2/C deposited
 under GenBank Accession Number AF007791, and to Xenopus
 laevis cement gland-specific proteins np77 and XAG-2
 deposited under GenBank Accession Numbers U82110 and
 AF025474, respectively"
 /codon_start=1
 /product="secreted cement gland protein XAG-2 homolog"
 /protein_id="AAC82614.1"
 /db_xref="GI:3779226"
 /translation="MERIPVSAFLILVALSTLARDTVKPKAKDTPKPKLPOTL
 SRGWDOLIMTQTYEALYKSKTNPLMIITHHIDEPHSQALKVFAENKEIOKLAE
 QEVLLNLYETTDKHLSPDQYVPRIMEVPSLVRADITGRYSNRLYAYEPADTALL
 LDNMKKALKLKTLEL"

BASE COUNT 367 a 231 c 210 g 269 t
 ORIGIN

alignment_scores:
 Quality: 194.00 Length: 194
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-266a-181 x AF038451 ..

Align seg 1/1 to: AF038451 from: 1 to: 1077

```

13 PROHISProSerArgArgLeuThrGlnGlyArgTrpValArgLysSer 29
|||||
2 CCGCATCTAGCCGCGACTCACACAGGCGAGTGGGTAGGAAATCCAG 51
29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaL 46
|||||
52 AGTTGCCATGGAGAAAATTCAGTGTGACATCTTCTCTGTTGCCCC 101
46 eSerTrpThrLeuAlaArgAspThrThrValLysProGlyAlaLys 62
|||||
102 TCTCTTACACTCTGGCCAGAGATACCAAGTCAACCTGGAGCCAAAAG 151
63 AspThrLysAspSerArgProLysLeuProGlnThrLysSerArgLys 79
|||||
152 GACACAAAGAGCTCTGACCCAAACAGTCCCAAGACCTCTCCAGAGTTG 201

```

79 pGlyAspGlnLeuIleThrPthrGlnThrTyrgLugLualLeuTyrlYss 96
|||||
202 GGGTACCACACTCATCTGACTCAGACATATGAGAAGCTCTATATAAT 251
96 erlysthrSerAsnlySProleumetlleleIHsHsLeuAspGluCys 112
|||||
252 CCAAGACACAGCAACAACTTGATGATTAATCATCTGACTTGATGAGTGC 301
113 ProHsSerGlnAlaLeuLysLysValPheAlaGlnAsnLysGluIleG1 129
|||||
302 CCACACAGCTCAGCTTAAAGAAAGTGTTCCTGAAAAATAAGAAATCCA 351
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgLutHrThra 146
|||||
352 GAAATGGCAGACAGAGTTTGTCTCTCAATCTGTTATGAAACACTG 401
146 sPlySHsLeuSerProAspGlyGlnTyValProArgIleMetPheVal 162
|||||
402 ACAAACACCTTCTCTGATGCGCAGTATGTCGCCAGGATTATGTGTT 451
163 AsPProSerLeuThrValArgAlaAspIlePthrGlyArgTySerAsnAr 179
|||||
452 GACCATCTGACAGATTAGAGCCGATATCATCTGAGATATTCAAATCG 501
179 gLeuTyAlaTyrgLupProAlaAspThrAlaLeuLeuAspAsnMetL 196
|||||
502 TCTCTATGCTTACGACCTGCAGATACAGCTCTGTGCTTGACAACATGA 551
196 ySLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
552 AGAAAGCTCTCAAGTGTCTGAAGACTGAATTG 583
seq_name: gb_pr:AF115926
seq_documentation_block:
LOCUS AF115926 1701 bp mRNA linear PRI 30-DEC-2001
DEFINITION Homo sapiens XAG-2 homolog long protein (HPC8) mRNA, complete cds.
ACCESSION AF115926
VERSION AF115926.1 GI:17998664
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Zhang,J.S. and Smith,D.I.
TITLE Identification of human homolog of XAG-2 over-expressed in tumors
JOURNAL Unpublished
2 (bases 1 to 1701)
AUTHORS Zhang,J.S. and Smith,D.I.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1998) Laboratory of Medicine and Pathology, Mayo
Clinic, 200 SW 1st Street, Rochester, MN 55905, USA
FEATURES
source
1..1701
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1701
/cell_type="prostate cancer"
/gene="HPC8"
58..585
/gene="HPC8"
/gene="HPC8"
/codon_start=1
/product="XAG-2 homolog long protein"
/protein_id="AA154870.1"
/db_xref="GI:17998665"
/translation="MERIPVSAPLLVALSYTLARPTTKPKAKKDTKDSRPKLPQTL
SRGMDQLMTQTYEBALYKSKTSNKPLMIHLLDCPSQALKVFAEKKEIQKLAE
QFVLNLVETTDKHLSPDGQYPRIMFVPSLTVRADITGRYSNRLYAVEPADTALL
LDNMKRAKLKLTREL"
BASE COUNT 557 a 352 c 339 g 453 t
ORIGIN

alignment_scores:
Quality: 194.00 Length: 194
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-266A-181 x AF115926
Align seq 1/1 to: AF115926 from: 1 to: 1701
13 ProHsProSerArgArgLeuThrGlnGlyArgThrValArgLysSerAr 29
|||||
1 CCGATCCTTACGCCGCCAGCTCTCACAAAGCGGTGGGTGAGAAATCCAG 50
29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaL 46
|||||
51 AGTGGCCATGGAGAAATTCAGATGTCAGCATTTCTGCTCTTGCGCC 100
46 euSerTyThrLeuAlaArgAspThrThraLysProGlyAlaLysLys 62
|||||
101 TCTCTACTACTCTGGCCAGAGATACACAGTCAACCTGAGCAAAAG 150
63 AsPThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
|||||
151 GACACAAAGACTCTCGACCCAACTGCCAGACCTCTCCAGAGGTG 200
79 pGlyAspGlnLeuIleThrPthrGlnThrTyrgLugLualLeuTyrlYss 96
|||||
201 GGGTGACCACTCATCTGACTCAGACATATGAAGAGCTCTATATAAT 250
96 erlysthrSerAsnlySProleumetlleleIHsHsLeuAspGluCys 112
|||||
251 CCAAGACACAGCAACAACTTGATGATTAATCATCTGACTTGATGAGTGC 300
113 ProHsSerGlnAlaLeuLysLysValPheAlaGlnAsnLysGluIleG1 129
|||||
301 CCACACAGCTCAGCTTAAAGAAAGTGTTCCTGAAAAATAAGAAATCCA 350
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgLutHrThra 146
|||||
351 GAAATGGCAGACAGATTGTCCTCTCAATCTGTTATGAAACACTG 400
146 sPlySHsLeuSerProAspGlyGlnTyValProArgIleMetPheVal 162
|||||
401 ACAAACACCTTCTCTGATGCGCAGTATGTCGCCAGGATTATGTGTT 450
163 AsPProSerLeuThrValArgAlaAspIlePthrGlyArgTySerAsnAr 179
|||||
451 GACCATCTGACAGATTAGAGCCGATATCATCTGAGATATTCAAATCG 500
179 gLeuTyAlaTyrgLupProAlaAspThrAlaLeuLeuAspAsnMetL 196
|||||
501 TCTCTATGCTTACGACCTGCAGATACAGCTCTGTGCTTGACAACATGA 550
196 ySLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
551 AGAAAGCTCTCAAGTGTCTGAAGACTGAATTG 582
seq_name: gb_pr:BC015503
seq_documentation_block:
LOCUS BC015503 836 bp mRNA linear PRI 29-OCT-2001
DEFINITION Homo sapiens, anterior gradient 2 (Xenopus laevis) homolog, clone
MGC:9112 IMAGE:3852448, mRNA, complete cds.
ACCESSION BC015503
VERSION BC015503.1 GI:15930124
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 836)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mdgaxil@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 20 Row: 0 Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6652811.

FEATURES
source
location/Qualifiers
1..836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="GI:15930125"
/clone="MGC:9112 IMAGE:3852448"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_65"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
57..584
/codon_start=1
/product="anterior gradient 2 (Xenopus laevis) homolog"
/product_id="AAH15503.1"
/db_xref="GI:15930125"
/translation="MEKTPVASFLLVALSYLTARDTPVKPKAKDTKDSRPKLPTPLSRGNGDQIMTQTEALYKSTSNKPLMIITHHDECHSQALKKVFENKEIOKLAE OFVLNLVYETTDKHLSPDGOYPRIMEVDSLVFADITGRYSNRLVAYEPADTALL LDNMKKALKLKTFL"

BASE COUNT 282 a 181 c 166 g 207 t
ORIGIN

alignment_scores:
Quality: 193.00 Length: 193
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x BC015503 ..

Align seg 1/1 to: BC015503 from: 1 to: 836

```

14 HisProSerArgLeuThrGlnGlyArgTrpValAlaGlySerArgVala 30
|||||
3 CATCTAGCGCGCGACTCACACAAAGCAGGTGGTGAGAAATCCAGAGT 52
|||||
30 AlaMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuS 47
|||||
53 TGGCATGGAGAAATTCAGTGTACAGATCTGCTCTGTGGCCCTCT 102
|||||
47 eertyrthrleualaargapthrthrValLysProGlyAlaLysLysasp 63
|||||
103 CTTAGCTCTGGCGCAGATACCAAGTCAACCTGGAGCCAAAAGGAG 132
|||||
64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgLysTrp 80
|||||
153 ACAAGAGACTCTGACCAAACTGCCCAAGACCTCTCCAGAGTGGGG 202

```

80 yaspGlnLeuIleTrpThrGlnThrTyrgLugLualalaLeuTyrlSerL 97
|||||
203 TGACCAACTCATCTGAGCTCAGACATATGAAGAAGCTCATATAATATCA 252
|||||
97 ysrHserasnLysProLeuMetIleIleHisLysLeuSpGluCysPro 113
|||||
253 AGACAAGCAACAACCCCTTGATGATATTATTCATCATCTGCGTGGTCCCA 302
|||||
114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGln 130
|||||
303 CACAGTCAAGCTTTAAAGAAAGTGTTCGTAAGAAATAAGAAATCCAGAA 352
|||||
130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValIlyrGluThrThrasp 147
|||||
353 ATGGCAGAGCAAGTTGTCTCTCAATCTGGTTATGAAACAACCTGACA 402
|||||
147 yHisLeuSerProAspGlyGlnTyTrValProArgIleMetPheValasp 163
|||||
403 AACACCTTTCCTGATGGCCAGTATGTCGCCAGATATATGTTGTTGAC 452
|||||
164 ProSerLeuThrValAlaAlaAspIleThrGlyArgTyrSerAsnArgle 180
|||||
453 CCATCTCTGACAGTGAAGCCGATATCAGTGAAGATATTCAAACCGCTCT 502
|||||
180 uTyraIaTyrgLuproAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
|||||
503 CTATGCTTACGAACCTGCAGATACAGCTCTGTGCTTGACACATGAAAGA 552
|||||
197 ySalaleuLysLeuLeuLysThrgLuleu 206
|||||
553 AACCTCTCAAGTTGCTGAAGACGTGAATTG 581

seq_name: gp_pat:ARI23828

seq_documentation_block:
LOCUS ARI23828 875 bp DNA Linear PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6171816.
ACCESSION ARI23828
VERSION ARI23828.1 GI:14109189
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 875)
AUTHORS Yu, G., Dillon, P. J., Ebner, R. and Endress, G. A.
TITLE Human XAG-1 polynucleotides and polypeptides
JOURNAL Patent: US 6171816-A 1 09-JAN-2001;
FEATURES
source
1..875
/organism="unknown"
BASE COUNT 283 a 189 c 180 g 223 t
ORIGIN

alignment_scores:
Quality: 192.00 Length: 192
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x ARI23828 ..

Align seg 1/1 to: ARI23828 from: 1 to: 875

```

15 ProSerArgArgLeuThrGlnGlyArgTrpValAlaGlySerArgVala 31
|||||
20 CTTAGCGCGCGACTCACACAGCAGGTGGTGAGAAATCCAGAGTTGC 69
|||||
31 aMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerT 48
|||||
70 CATGGAGAAATTCAGTGTACAGATCTGCTCTGTGGCCCTCTCCCT 119

```

Align seg 1/1 to: AX341084 from: 1 to: 557

```

48 yrrhrleualarqasprthrValLysProGlyAlaLysLysAspThr 64
120 AACCTCTGGCAGAGATACACAGTCAACCTGGAGCCAAAAGACACA 169
65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
170 AAGGACTCTGCACCAACTGCCACAGCCCTCTCCAGAGGTGGGGGA 219
81 pGlnLeuIleTrpThrGlnThrTyrgLuglUAlaLeuTyrlsSerLysT 98
220 CCAACTCATCTGGACTCAGACATATGAAGACCTCTATATTAATCAAGA 269
98 hrSerAsnLysProLeuMetIleIleHisLysLeuAspGlyCysProHis 114
270 CAAGCAACAAACCCCTTGATGATTATTCATCACTTGATGAGGCCACAC 319
115 SerGlnAlaLeuLysValPheAlaGluAsnLysGluIleGlnLysLe 131
320 AGTCAAGCTTTAAAGAAAGTGTGGCTGAATAAAGAAATCCAGAAATTT 369
131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrgLuglUAlaLeuTyrls 148
370 GGCAGACGAGTGTCTCTCTCAATCTGTTATGAACAACACTGCAAC 419
148 lSLeuSerProAspGlyGlnTyrgValProArgIleMetPheValAspPro 164
420 ACCCTTTCCTGATGGCCAGATATGCCAGGATATGTTGTTGACCACA 469
165 SerLeuThrValArgAlaAspIleThrGlyArgTyrgSerAsnArgLeuTy 181
470 TCTCTGACAGTTAGAGCCGATATCTGGAAGATATTCAAATCGCTCTTA 519
181 rAlaTyrgLuproAlaAspThrAlaLeuLeuAspAsnMetLysLysA 198
520 TGCTTACGACACTGCAGATACAGCTCTGTGCTTGACAAACATGAGAAAG 569
198 lAlaLeuLysLeuLeuLysThrGluLeu 206
570 CTCCTCAAGTTGCTGAAGACTGAAATG 595

```

seq_name: gb_pat:AX341084

seq_documentation_block: 557 bp DNA linear PAT 10-JAN-2002

LOCUS AX341084
 DEFINITION Sequence 1331 from Patent WO0196388.
 ACCESSION AX341084
 VERSION AX341084.1 GI:18137066
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Jiang,Y., Harlocker,S.L. and Secrist,H.
 TITLE Compositions and methods for the therapy and diagnosis of colon cancer

JOURNAL Patent: WO 0196388-A 1331 20-DEC-2001;
 CORIXA CORPORATION (US)
 FEATURES
 source 1..557
 location/Qualifiers

BASE COUNT 169 a 143 c 113 g 130 t 2 others
 ORIGIN
 alignment_scores:
 Quality: 184.00 Length: 184
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX341084

Align seg 1/1 to: AX341084 from: 1 to: 557

```

13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArg 29
1 CGGATCTCTAGCCGCCGACTCACACAGGCGAGTGGGAGAGAAATCCAG 50
29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuValAlaL 46
51 AGTGGCATGGAGAAATTCAGATGTCCAGCATTTCTTGCTCTTGGGCC 100
46 euSerTyrrhrleualarqasprthrValLysProGlyAlaLysLys 62
101 TCTCTTACACTCTGGCCAGATATCCACAGTCAACCGGAGGCCAAAAAG 150
63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
151 GACACAAAGAGACTCTGCACCAACTGCCACAGACCTCTCCAGAGGTTG 200
79 pGlyAspGlnLeuIleTrpThrGlnThrTyrgLuglUAlaLeuTyrls 96
201 GGGTGACCACTCATCTGACTCAGACATATGAAGAAGCTCTATATAAT 250
96 eLysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGlyCys 112
251 CCAAGACACAGCAACAAACCTTGATGATTATTCATCACTTGATGAGATG 300
113 ProHisSerGlnAlaLeuLysValPheAlaGluAsnLysGluIleGln 129
301 CCACACAGTCAAGCTTTAAAGAAAGTGTGCTGAAATAAAGAAATCCA 350
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgLuglUAla 146
351 GAAATGGCAGAGCAGTTGTCTCTCAATCTGTTATGAACAACACTG 400
146 sPlyHisLysLeuSerProAspGlyGlnTyrgValProArgIleMetPheVal 162
401 ACAAAACACCTTCTCTGATGGCCAGATATGCCAGGATATGTTGTT 450
163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrgSerArg 179
451 GACCCATCTCTGACAGCTTAGAGCCGATATCACTGGAATATTCAAATCG 500
179 gLeuTyrrhrleualarqasprthrAlaLeuLeuAspAsnMetL 196
501 TCTCTATGCTTACGAACTGCAGATACAGCTCTGTGCTTGACAAACATGA 550
196 ys 196
551 AG 552

```

seq_name: gb_pat:AX231601

seq_documentation_block: 543 bp DNA linear PAT 11-SEP-2001

LOCUS AX231601
 DEFINITION Sequence 7 from Patent WO0163290.
 ACCESSION AX231601
 VERSION AX231601.1 GI:15592500
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Boyd,R.S., Stamps,A.C., Terrett,J.A. and Tyson,R.L.
 TITLE Bcnp-7 as marker for diagnosis of breast cancer

JOURNAL Patent: WO 0163290-A 7 30-AUG-2001;
 Oxford GlycoSciences (UK) Limited (GB)
 FEATURES
 source 1..543
 location/Qualifiers

alignment_block:
 US-09-674-266A-181 x AX231601

```

/note="unnamed protein product"
/codon_start=1
/protein_id="CAC69703.1"
/db_xref="GI:15592501"
/translation="MEKIPVSAFLLVALSTTLARDTTPKGAANKDTPKSPKLPDTL
SRGWDQILMTQTYEEALYKSTSKNPIMTIHHDECHSQAALKVFAENKEIQKLA
QFVLNLVYETTDKHLSPDQGYVPRIMEVDSLTVRADITGRYSNRLYAYBPADTALL
LDNMKALKLKTTEL"
BASE COUNT      170 a      135 c      107 g      131 t
ORIGIN

alignment_scores:
    Quality: 178.00      Length: 178
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x AX231601 ..

Align seg 1/1 to: AX231601 from: 1 to: 543

29  ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAla 45
|||||
2  AGAGTTGCCATGGAGAAAATCCAGTGTGACATTTCTGCTCTGTTGGC 51
|||||
45  aleuSerTyrThrLeuAlaArgAspThrThrValLysProGlyValAla 62
|||||
52  CCCTCTCTACACTGTGGCCAGAGATACACAGTCACAAACGAGCCAAAA 101
|||||
62  ysasprThrLysaspSerArgProLysLeuProGlnThrLeuSerArgGly 78
|||||
102 AGAGACAAAGAGACTCTCGACCCAACTGCCAGACCCCTCTCCAGAGGT 151
|||||
79  TrpGlyAspGlnLeuIleTyrThrGlnThrTyrGluGluAlaLeuTyr 95
|||||
152 TGGGGTGACCACTCATCTGACACACATGATGAAGAAGCTCATATTA 201
|||||
95  sserLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 112
|||||
202 ATCCAAAGACAGCAACAAACCTTGATGATTAATCATCTGATGATGAT 251
|||||
112 yspProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGlu 128
|||||
252 GCCCACACAGCTCAAGCTTTAAAGAAAGTGTGCTGAAATTAAGAATC 301
|||||
129 GlnLysLeuAlaGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
|||||
302 CAGAAATTGGCAGAGCAGTTGTCTCTCATCTGTTATGAAACAAC 351
|||||
145 raapLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 162
|||||
352 TGACAAACACCTTTCTCTGATGCGCCAGTATGTCGCCAGATATGTTG 401
|||||
162 alaaspProSerIleThrValArgAlaAspIleThrGlyArgTyrSer 178
|||||
402 TTGACCCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTA 451
|||||
179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuAspAsme 195
|||||
452 CGTCTCATGCTTACGAAACCTGCGAGATACAGCTCTGTTGCTGCAAC 501
|||||
195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
502 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 535
|||||

seq_name: gb_pr:AF007791

seq_documentation_block:
LOCUS      AF007791          838 bp      mRNA      linear      PRI 14-NOV-1998
DEFINITION Homo sapiens secreted cement gland protein XAG-2 homolog (hag-2/c)
            mRNA, complete cds.
ACCESSION  AF007791

```

```

VERSION      AF007791.1  GI:3779196
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 838)
AUTHORS     Thompson, D.A. and Miegel, R.J.
TITLE       hag-2, the human homologue of the Xenopus laevis cement gland gene
            XAG-2, is coexpressed with estrogen receptor in breast cancer cell
            lines
JOURNAL     Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)
MEDLINE     9909231
REFERENCE   2 (bases 1 to 838)
AUTHORS     Thompson, D.A.
TITLE       Direct Submission
            Submitted (10-JUN-1997) Surgery, Stanford, 1201 Welch Road, MSIS
            Building, Room P228, Stanford, CA 94305, USA
FEATURES
            source
            1..838
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /cell_line="MCF7; estrogen receptor-expressing"
                /cell_type="adenocarcinoma"
                /tissue_type="breast"
                1..838
                /gene="hag-2/c"
                11..538
                /gene="hag-2/c"
                /note="similar to Homo sapiens protein hag-2/R deposited
                in GenBank Accession Number AF038451, and to Xenopus
                laevis cement gland proteins np77 deposited in GenBank
                Accession Number U82110 and XAG-2 deposited in GenBank
                Accession Number AF025474"
                /codon_start=1
                /product="secreted cement gland protein XAG-2 homolog"
                /protein_id="AAC77358.1"
                /db_xref="GI:3779197"
                /translation="MEKIPVSAFLLVALSTTLARDTTPKGAANKDTPKSPKLPDTL
                SRGWDQILMTQTYEEALYKSTSKNPIMTIHHDECHSQAALKVFAENKEIQKLA
                EQLNLVYETTDKHLSPDQGYVPRIMEVDSLTVRADITGRYSNRLYAYBPADTALL
                LDNMKALKLKTTEL"
BASE COUNT      291 a      170 c      160 g      217 t
ORIGIN

alignment_scores:
    Quality: 178.00      Length: 178
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x AF007791 ..

Align seg 1/1 to: AF007791 from: 1 to: 838

29  ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAla 45
|||||
2  AGAGTTGCCATGGAGAAAATCCAGTGTGACATTTCTGCTCTGTTGGC 51
|||||
45  aleuSerTyrThrLeuAlaArgAspThrThrValLysProGlyValAla 62
|||||
52  CCCTCTCTACACTGTGGCCAGAGATACCAAGTCACAACTGGAGCCAAA 101
|||||
62  ysasprThrLysaspSerArgProLysLeuProGlnThrLeuSerArgGly 78
|||||
102 AGAGACAAAGAGACTCTCGACCCAACTGCCAGACCCCTCCAGAGGT 151
|||||
79  TrpGlyAspGlnLeuIleTyrThrGlnThrTyrGluGluAlaLeuTyr 95
|||||
152 TGGGGTGACCACTCATCTGACACACATGATGAAGAAGCTCATATTA 201
|||||
95  sserLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 112
|||||

```

```

|||||
202 ATCCAGACAAACCAACCTTGATGATATTCATCAGTGAAGT 251
112 yspHhisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluLeu 128
252 GCCACACAGCTACAGCTTTAAAGAAAGTGTTCGAAATAAAGAAATC 301
129 GlnLysLeuAlaGluGlnPheValLeuAsnLeuValTyrGluThrTh 145
302 CAGAAATTTGGCAGACAGTGTTCCTCCATCGTGTATGAAACAAAC 351
145 rasplyshLysLeuSerProaspGlyGlnTyrValProarglleMetPhe 162
332 TGACAAACACCTTCTCCATGAGCCAGATGTCCTCCAGGATTATGTTG 401
162 alasproSerLeuThrValAlaArgAlaAspIleThrGlyArgTyrSerAsn 178
402 TTGACACCATCTTGACAGTTAGACCGCATATCAGTGAAGATATTCAAAT 451
179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuAspAsnMe 195
452 CGTCTATAGCTTACGAACCTGCAGATACAGCTGTGCTTGACAAACAT 501
195 tLysLysAlaLeuLysLeuLysThrGluLeu 206
502 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATG 535
seq_name: gb_pat:AR177367

```

seq_documentation_block:

```

LOCUS AR177367 1689 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 61 from patent US 6312922.
ACCESSION AR177367
VERSION AR177367.1 GI:17919722
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1689)
AUTHORS Edwards,J.-B.Dumas, Mline., Duclet,A. and Bougueleret,L.
TITLE Complementary DNAs
JOURNAL Patent: US 6312922-A 61 06-NOV-2001;
FEATURES
source
1..1689
location/Qualifiers
BASE COUNT 552 a 350 c 335 g 452 t
ORIGIN

```

alignment_scores:

```

Quality: 178.00 Length: 178
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

```

US-09-674-266A-181 x AR177367

```

```

Align seg 1/1 to: AR177367 from: 1 to: 1689

```

```

29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
|||||
42 AGAGTTGCCATGAGAAATTCACAGTCACGATTCCTGCTGTCGTCG 91
45 aleuSerTyrThrLeuAlaArgAspThrThValLysProGlyAlaLysL 62
|||||
92 CCTCTCCATCACTTGCCAGAGATACCACTCAAACTGAGCCCAAAA 141
62 ysaSPThrLysAspSerArgProLysLeuProGlnThrLeuSerArgly 78
|||||
142 AGGACACAAGAGACTCGACCAACGCCCCAGACCCCTCCAGAGGT 191
79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrly 95
|||||

```

```

192 TGGGCTGACCAACTCATCTGACACTGACATATGAAAGACTCTATATA 241
95 sSerLysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluC 112
242 ATCCAGACAAACCAACCTTGATGATATTCATCAGTGAAGT 291
112 yspHhisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluLeu 128
292 GCCACACAGCTACAGCTTTAAAGAAAGTGTTCGAAATAAAGAAATC 341
129 GlnLysLeuAlaGluGlnPheValLeuAsnLeuValTyrGluThrTh 145
342 CAGAAATTTGGCAGACAGTGTTCCTCCATGAGCCAGATGTCCTCCAG 391
145 rasplyshLysLeuSerProaspGlyGlnTyrValProarglleMetPhe 162
392 TGACAAACACCTTCTCCATGAGCCAGATGTCCTCCAGGATTATGTTG 441
162 alasproSerLeuThrValAlaArgAlaAspIleThrGlyArgTyrSerAsn 178
442 TTGACACCATCTTGACAGTTAGACCGCATATCAGTGAAGATATTCAAAT 491
179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuAspAsnMe 195
492 CGTCTATAGCTTACGAACCTGCAGATACAGCTGTGCTTGACAAACAT 541
195 tLysLysAlaLeuLysLeuLysThrGluLeu 206
542 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATG 575
seq_name: gb_pat:AX341234

```

seq_documentation_block:

```

LOCUS AX341234 603 bp DNA linear PAT 10-JAN-2002
DEFINITION Sequence 1481 from Patent WO0196388.
ACCESSION AX341234
VERSION AX341234.1 GI:18137216
KEYWORDS
SOURCE
ORGANISM human.
REFERENCE
1 (sites)
AUTHORS Jiang,Y., Harlocker,S.L. and Secrist,H.
TITLE Compositions and methods for the therapy and diagnosis of colon
cancer
JOURNAL Patent: WO 0196388-A 1481 20-DEC-2001;
FEATURES
CORIXA CORPORATION (US)
source
1..603
location/Qualifiers
BASE COUNT 187 a 153 c 116 g 144 t 3 others
ORIGIN

```

alignment_scores:

```

Quality: 177.00 Length: 177
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

```

US-09-674-266A-181 x AX341234

```

```

Align seg 1/1 to: AX341234 from: 1 to: 603

```

```

29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
|||||
30 AGAGTTGCCATGAGAAATTCACAGTCACGATTCCTGCTGTCGTCG 79
45 aleuSerTyrThrLeuAlaArgAspThrThValLysProGlyAlaLysL 62
|||||
80 CCTCTCCATCACTTGCCAGAGATACCACTCAAACTGAGCCCAAAA 129
|||||

```

```

62 ysaSprhlysaSpSerArgProlyseupProGlnThrleuSerArgly 78
   |||
130 AGACACAAAGAGACTCTCGACCAACTGCCAGACCTCTCCAGAGGT 179
   |||
79 TrpGlyAspGlnleuIleTrpThrGlnThrTyrgluGlnAlaLeuTyrl 95
   |||
180 TGGGGTGACCAACTCATCTGACTGCACATATGAAGAACCTCTATATTA 229
   |||
95 sSerlysthrSerAsnlyseProleuMetIleIleHisHisleuAspGlu 112
   |||
230 ATCCAGACAAAGACAAACCTTGATGATTAATTCACCTTGATGAGT 279
   |||
112 ySProHISserGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
   |||
280 GCCCAGACAGTCAAGCTTTAAAGAAAGTGTTCGTAATAAAGAAATC 329
   |||
129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgluThr 145
   |||
330 CAGAAATTTGGCAGAGCAGTTGTCTCTCCATCTGGTTATGAACAC 379
   |||
145 rasplyshISleuSerProaspGlyGlnTyraProArgIleMetPhe 162
   |||
380 TGACAAACACCTTTCCTGATGCCAGATATGCCAGGATTAATGTTTG 429
   |||
162 alaspproSerleuThrValArgAlaAspIleThrGlyArgTySerAsn 178
   |||
430 TTACACCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATTAATCAAT 479
   |||
179 ArgLeuTyraLAtyrgluProAlaAspThrAlaLeuLeuAspAsn 195
   |||
480 CGTCTCATCTTACGAACTCGAGATACAGCTCTGTCTTGACAAAT 529
   |||
195 tlyslsAlaLeuLysleuLeuLysThrGlu 205
   |||
530 GAAGAAAGCTCTCAAGTCTGGAAGACTGAA 560

```

seq_name: gb_pat:AX341953

seq_documentation_block:

```

LOCUS AX341953 525 bp DNA linear PAT 10-JAN-2002
DEFINITION Sequence 2200 from Patent WO0196388.
ACCESSION AX341953
VERSION AX341953.1 GI:18137935
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Jiang,Y., Harlocker,S.L. and Secrist,H.
TITLE Compositions and methods for the therapy and diagnosis of colon
cancer
JOURNAL Patent: WO 0196388-A 2200 20-DEC-2001;
AUTHORS CORIXA CORPORATION (US)
TITLE Location/Qualifiers
FEATURES
source 1..525
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 156 a 140 c 102 g 125 t 2 others
ORIGIN

```

alignment_scores:

```

Quality: 161.00 Length: 161
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

US-09-674-266A-181 x AX341953 ..

Align seg 1/1 to: AX341953 from: 1 to: 525

```

29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
   |||
31 AGAGTTGCCATGAGAAAATTCCAGTGTGCAGATTTCTGCTTGTCGC 80
   |||
45 aleuSerTyThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
   |||
81 CCTCTCCTACACTGTCGCCAGAGATACACAGTCAAACTCGAGCCAAA 130
   |||
62 ysaSprhlysaSpSerArgProlyseupProGlnThrleuSerArgly 78
   |||
131 AGACACAAAGAGACTCTCGACCAACTGCCAGACCTCTCCAGAGGT 180
   |||
79 TrpGlyAspGlnleuIleTrpThrGlnThrTyrgluGlnAlaLeuTyrl 95
   |||
181 TGGGGTGACCAACTCATCTGACTGCACATATGAAGAACCTCTATATTA 230
   |||
95 sSerlysthrSerAsnlyseProleuMetIleIleHisHisleuAspGlu 112
   |||
231 ATCCAGACAAAGACAAACCTTGATGATTAATTCACCTTGATGAGT 280
   |||
112 ySProHISserGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
   |||
281 GCCCAGACAGTCAAGCTTTAAAGAAAGTGTTCGTAATAAAGAAATC 330
   |||
129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgluThr 145
   |||
331 CAGAAATTTGGCAGAGCAGTTGTCTCTCCATCTGGTTATGAACAC 380
   |||
145 rasplyshISleuSerProaspGlyGlnTyraProArgIleMetPhe 162
   |||
381 TGACAAACACCTTTCCTGATGCCAGATATGCCAGGATTAATGTTTG 430
   |||
162 alaspproSerleuThrValArgAlaAspIleThrGlyArgTySerAsn 178
   |||
431 TTACACCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATTAATCAAC 480
   |||
179 ArgLeuTyraLAtyrgluProAlaAspThrAla 189
   |||
481 CGTCTCATCTTACGAACTCGAGATACAGCT 513

```

seq_name: gb_pat:AX351468

seq_documentation_block:

```

LOCUS AX351468 699 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 215 from Patent WO0196390.
ACCESSION AX351468
VERSION AX351468.1 GI:18616815
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Jiang,Y., Hepler,W.T., Clapper,J.D., Wang,A. and Secrist,H.
TITLE Compositions and methods for the therapy and diagnosis of colon
cancer
JOURNAL Patent: WO 0196390-A 215 20-DEC-2001;
AUTHORS CORIXA CORPORATION (US)
TITLE Location/Qualifiers
FEATURES
source 1..699
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 181 a 139 c 154 g 224 t 1 others
ORIGIN

```

alignment_scores:

```

Quality: 160.00 Length: 160
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

US-09-674-266A-181 x AX351468/rev ..

Align seg 1/1 to reverse of: AX067336 from: 1 to: 699

```

47 SerTyrThrLeuAlaArgAspThrThrValLysProGluAlaLysLysAs 63
|||||
658 TCCTACACTCTGGCCAGATACCACTCAAACTGGAGCCAAAAGA 609
63 pThrLysAspSerArgProLysLeuProGluThrLeuSerArgLysTrp 80
|||||
608 CACAAGAGACTCTGACCCAAAGTCCCGACAGCCCTCTCCAGAGTTGGG 559
80 LysArgGlnLeuIleTyrThrGlnThrTyrGluGluAlaLeuTyrLysSer 96
|||||
558 GTGACCAACTCATCTGACTCAGACATATGAAAGAGCTTATATTAATCC 509
97 LysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluCysPr 113
|||||
508 AAGACAAAGCAAAACCTTGATGTTTATTCATCTCGATGAGTGCC 459
113 OHISerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnL 130
|||||
458 ACACAGCTCAAGCTTTAAAGAAAGTGTGCTGAATAATAAGAAATCCAGA 409
130 LysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAsp 146
|||||
408 AATTGGCAGAGCAGTTTGTCTCTCAATCTGTTTATGAAACAACCTGAC 359
147 LysHisLeuSerProAspGluGlnTyrValProArgIleMetPheValAs 163
|||||
358 AAACACCTTTCTCCGATGGCCAGATCTCCCGAGATTATGTTGTGA 309
163 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgL 180
|||||
308 CCCATCTCTGACAGTTAGACCGATATCAGTGAAGATTTCAATCGTC 259
180 eUTyrAlaTyrGluProAlaAspThrAlaLeuLeuAspAsnMetLys 196
|||||
258 TCTAGCTTAGCAACCTGACAGTACAGCTCTGTGCTTGACACAACATGAA 209
197 LysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
208 AAGCTCTCAAGTTCTGTAAGACTGAATTG 179

```

seq_name: gb_pat:AX067336

seq_documentation_block:

LOCUS AX067336 778 bp DNA linear PAT 24-JAN-2001

DEFINITION Sequence 40 from Patent WO0078960.

ACCESSION AX067336

VERSION AX067336.1 GI:12544960

KEYWORDS

SOURCE

ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE Yugi,J. and Mitcham,J.L.

AUTHORS 1 (bases 1 to 778)
TITLE Compositions and methods for the therapy and diagnosis of breast cancer

JOURNAL Patent: WO 0078960-A 40 28-DEC-2000;

FEATURES CORIXA CORPORATION (US)

source 1..778
Location/Qualifiers

misc_feature 1..778
/db_xref="taxon:9606"

BASE COUNT 245 a 175 c 158 g 193 t 7 others

ORIGIN

alignment_scores:

Quality: 134.00 Length: 134

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266A-181 x AX067336 ..

Align seg 1/1 to: AX067336 from: 1 to: 778

```

54 ThrThrValLysProGluAlaLysLysAspThrLysAspSerArgProLys 70
|||||
125 ACCACAGTCAAAAGCTCTATATTAATCCAAAGACCAAGAGACTCTCGACCAA 174
70 sLeuProGlnThrLeuSerArgLysTyrGluAspGlnLeuIleTyrThr 87
|||||
175 ACTGCCAGACCTCTCCAGAGGTTGGGTGACCAACTCATCTGGACTC 224
87 LntTyrGluGluAlaLeuTyrLysSerLysThrSerAsnLysProLeu 103
|||||
225 AGACATATGAAGAAGCTCTATATTAATCCAAAGACCAACCAACCTTG 274
104 MetIleIleHisLysLeuAspGluCysProHisSerGlnAlaLeuLysL 120
|||||
275 ATGATTATTATCATCACTTGATGATGAGTGCCACACAGCTCAACCTTTAAAGAA 324
120 sValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGlnPheValL 137
|||||
325 AGTGTTCCTGAAATAATAAGAAATCCAGAAATTTGGCAGAGCAGTTGTGCC 374
137 eUeAsnLeuValTyrGluThrThrAspLysHisLysSerProAspGly 153
|||||
375 TCCTCAATCTGTTTATGAACCAACTGACAAACACTTTCTCCGATGGCC 424
154 GlnTyrValProArgIleMetPheValAspProSerLeuThrValArgAl 170
|||||
425 CAGTATGTCCCGACGATATGTTGTTGACCATCTCGACAGTTAGAGC 474
170 aAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGluProAla 187
|||||
475 CGATATCACTGGAAAGATATTCAAAATCGTCTATGCTTACGAAACCTGCAG 524
187 sp 187
||
525 AT 526

```

OM of: US-09-674-266a-181 to: N_Geneseq_032802:* out_format : pfs
Date: Oct 8, 2002 11:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODE=firmet-p2n.model -DEV=xlh  
-Q=/cgn2_1/USPTO.spool/US09674266/runcat.08102002_093410_27035/app.query.fasta_1.271  
-De=N_Geneseq_032802 -OFMT=fastap -SUFFIX=olip2n.rng  
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPEXT=0.000  
-LOOPEXT=0.000 -GAPOP=4.500 -GAPEXT=0.050 -GAPOP=60.000  
-GAPEXT=60.000 -GAPOP=6.000 -GAPEXT=7.000 -GAPOP=60.000  
-GAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR_SCORE=quality -THR_MIN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09674266.@CGN1_1.0 -NCPU=6 -ICPU=3 -LONGLOC  
-DEV_TIME=120 -WARN_TIME=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-674-266a-181
Query length: 206
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 211.730000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	! Documentation	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA52865 +			206.00	3721.06	6.5e-199	10
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV29047 +			195.00	3522.39	7.5e-188	86
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA538380 +			195.00	3521.94	8.0e-188	92
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH35026 +			194.00	3503.59	8.4e-187	9
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ77486 +			194.00	3503.01	9.0e-187	10
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV19155 +			192.00	3467.80	8.2e-185	87
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV63314 +			192.00	3467.80	8.2e-185	87
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV59320 +			191.00	3449.80	8.5e-184	88
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA513480 +			178.00	3216.97	7.7e-171	5
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ08083 +			178.00	3208.49	2.3e-170	12
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAZ29048 +			175.00	3162.71	8.1e-168	52
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC00115 +			146.00	2636.59	1.6e-138	46
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ44884 +			134.00	2414.74	3.7e-126	7
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV7903 +			133.00	2401.37	2.1e-125	40
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI28641 +			119.00	2147.18	3.0e-111	39
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV11145 +			118.00	2127.01	4.0e-110	52
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ22826 +			112.00	2020.60	3.4e-104	3
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV19619 +			107.00	1929.35	4.1e-99	38
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV19188 +			107.00	1929.35	4.1e-99	38
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV63347 +			103.00	1858.04	3.8e-95	3
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ47160 +			103.00	1858.04	3.8e-95	3
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI17730 +			103.00	1858.04	3.8e-95	3
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI55526 +			103.00	1858.31	6.1e-95	8
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAZ24578 +			102.00	1833.12	9.3e-94	79
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ5817 +			102.00	1833.12	9.3e-94	79
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAK09584 +			101.00	1822.03	3.9e-93	30
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI20462 +			98.00	1760.39	1.0e-76	9
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI23173 +			85.00	1522.96	1.8e-76	9
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAK15919 +			84.00	1512.88	6.4e-76	31
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI10039 +			76.00	1365.25	1.1e-67	4
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI11564 +			76.00	1361.48	1.7e-67	7
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI17753 +			66.00	1184.04	1.3e-57	40
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI28491 +			66.00	1184.04	1.3e-57	4
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ40846 +			62.00	1104.99	3.4e-53	94
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ5978 +			60.00	1074.96	1.6e-51	30
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV7841 +			58.00	1040.81	1.3e-49	30
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAI28579 +			58.00	1039.81	1.4e-49	34

/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AA558854 + 56.00 1007.49 9.1e-48
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV19189 + 53.00 948.35 1.8e-44
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV63348 + 53.00 948.35 1.8e-44
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV61263 + 52.00 936.35 8.1e-44
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV58679 + 52.00 936.35 8.1e-44
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAO6442 + 52.00 936.55 8.1e-44
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA563650 + 52.00 936.55 8.1e-44

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA52865

seq_documentation_block:

ID AA52865 standard: cDNA: 1020 BP.

AC AA52865;

DT 14-MAR-2000 (first entry)

DE Human prostate tumor cDNA library derived EST fragment #8.

KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;

KW treatment; ds.

OS Homo sapiens.

PN DE19820190-A1.

PD 04-NOV-1999.

PF 28-APR-1998; 98DE-1020190.

PR 28-APR-1998; 98DE-1020190.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR WPI; 1999-621386/54.

DR P-PSDB; AAV73835, AAV73836, AAV73837.

PT New human nucleic acid sequences from pancreatic tumors, and related

PS proteins -

PS Claim 2; Page 188; 502pp; German.

CC This invention describes novel polypeptides and their encoding nucleic

CC acids derived from human pancreatic tumor tissue which have cytostatic

CC activity. The sequences are also useful in producing pharmaceutical

CC compositions for treatment of pancreatic tumors. AA52858-253014

CC represent expressed sequence tag (EST) fragments derived from a human

CC pancreatic tumor cDNA library and which encode the proteins represented

CC in AAV73814-Y74252.

CC Sequence 1020 BP; 341 A; 214 C; 218 G; 247 T; 0 other;

CC alignment_scores: 206 00 Length: 206

CC Quality: 1.000 Ratio: 0 Gaps: 0

CC Percent Similarity: 100.000 Percent Identity: 100.000

CC alignment_block:

CC US-09-674-266a-181 x AA52865 ..

CC Align seg 1/1 to: AA52865 from: 1 to: 1020

CC 1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17

CC 57 AGACTCAGCTGCTGCTGACACACACGAGGCTGACACCGATCTTACGCC 106

CC 17 gargleuThrGlnGlyArgTrpValArgIysSerArgValAlaMetGluL 34

CC 107 CCGACTTCACACAGGAGGAGGCTGAGGAATCATCAGAGTTGCATGAGAGA 156


```
413 CTGCAACAACACCTTTCCTGATGGCAGATAGTCCAGATTATGTT 462
162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
463 GTTGACCACTCTCTGACAGATTAGAGCCGATATCATCTGGAAGATATCA 512
178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsn 195
513 CCGTCTCTATGCTTACGACACCTGACATACAGCTCTGTTGCTTGACACA 562
195 eTlYsAlaLeuLysIleuLeuLysThrGluLeu 206
563 TGAAGAAAGCTCTCAAGTTGCTGAAAGCTGAATTG 597
seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT: AAC58380
seq_documentation_block:
ID AAC58380 standard; cDNA; 920 BP.
XX
AC AAC58380;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human PRO1030 nucleotide sequence SEQ ID NO:39.
XX
KM Human: tumour; diagnosis: neoplastic disease; neoplastic cell growth;
KM proliferation; tumorigenesis; identification; cancer; cytostatic;
KM neurotropic; neuroprotective; antiinflammatory; immunosuppressive;
KM immunostimulant; antiangiogenic; leukemia; lymphoid malignancy;
KM neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KM hypochlamic disorder; glandular disorder; macrophagal disorder;
KM epithelial disorder; stromal disorder; blastocoealic disorder;
KM inflammatory disorder; immunologic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200053755-A2.
XX
PD 14-SEP-2000.
XX
XX
PF 06-JAN-2000; 2000MO-US00376.
XX
PR 08-MAR-1999; 99MO-US05028.
PR 02-JUN-1999; 99MO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 26-JUL-1999; 99US-0145698.
PR 30-NOV-1999; 99MO-US28313.
PR 20-DEC-1999; 99MO-US30911.
PR 05-JAN-2000; 2000MO-US00219.
XX
PA (GETH ) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;
PI Watanabe CK, Wood WI;
XX
DR WPI: 2000-572270/53.
XX
P-PSDB: AAB24070.
XX
PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
XX treatment, diagnosis and prevention of cancer -
XX
PS Claim 50; Fig 27; 286pp; English.
XX
CC The present invention describes an isolated antibody that binds to
CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC growth. The PRO polypeptides and nucleotides are useful in the
CC treatment, diagnosis and prevention of cancer. The antibodies and other
CC anti-tumour compounds may be used to treat various conditions, including
```

```
CC those characterised by overexpression and/or activation of the amplified
CC PRO genes. Exemplary conditions or disorders to be treated with such
CC antibodies and other compounds include benign or malignant tumours
CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC leukaemias and lymphoid malignancies, other disorders such as neuronal,
CC glial, astrocytal, hypochlamic and other glandular, macrophagal,
CC epithelial, stromal and blastocoealic disorders, and inflammatory,
CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC primers and hybridisation probes used in the isolation of the human PRO
CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.
XX
SQ Sequence 920 BP; 296 A; 198 C; 195 G; 231 T; 0 other;
XX
alignment_scores:
Quality: 195.00 Length: 195
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-266A-181 x AAC58380 ..
Align seg 1/1 to: AAC58380 from: 1. to: 920
12 GlyProHisProSerArgArgLeuThrGlnGlyArgTyrValArgLysSe 28
|||||
88 GGACCGCATCTCTACCGCCGCCGACTCACACAAAGCGAGTGGGAGGAATC 137
|||||
28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAla 45
|||||
138 CAGAGTTGCCATGAGAAATTCAGTCTCAGCATCTTCTGCTTGCG 187
|||||
45 lAlaSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
|||||
188 CCTCTCTACACTCTGGCCAGAGATACCAAGTCAAACTCGAGCCAA 237
|||||
238 AAGACACAAAGAGACTCTCGACCAAACTGCCCAAGCTCTCCAGAG 287
|||||
62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgG 78
|||||
188 CCTCTCTACACTCTGGCCAGAGATACCAAGTCAAACTCGAGCCAA 237
|||||
78 YTrpGlyAspGlnLeuIleTrrpThrGlnThrTyrGluAlaLeuTyrL 95
|||||
288 TTGGGGTGCACCACTCATCTGACTCAGACATATGAACAACTCTATATA 337
|||||
95 YSerLysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGlu 111
|||||
338 AATCCAAAGACAAACAAACCTTGATGATTATTCATCACTTGATGAG 387
|||||
112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluI 128
|||||
388 TGCCACACAGCTCAAGCTTAAAGAAAGTGTGGTGAATAAAGAAAT 437
|||||
128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluTrp 145
|||||
438 CCAGAAATTTGGCAGAGCAGTTTGTCTCTCAATCTGTTATGAAACAA 487
|||||
145 hAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
|||||
488 CTGACAAACACCTTCTCTCATGCGCATGTCCTCCAGGATTTATGTTT 537
|||||
162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
538 GTTGACCACTCTCTGACAGATTAGAGCCGATATCATCTGGAAGATATCA 587
|||||
178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsn 195
588 TCGTCTCTATGCTTACGACACCTGACATACAGCTCTGTTGCTTGACACA 637
|||||
195 eTlYsAlaLeuLysIleuLeuLysThrGluLeu 206
```

|||||
638 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 672

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH35026

seq_documentation_block:
ID AAH35026 standard; cDNA: 943 BP.

XX AAH35026:
XX
XX
XX 03-SEP-2001 (first entry)

DE Human colon cancer antigen encoding cDNA SEQ ID NO:2108.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma; ss.

XX Homo sapiens.

PN W020012920-A2.

PD 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 9905-0157137.

PR 03-NOV-1999; 9905-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

DR P-PSDB; AAG75621.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 1; Page 3593; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 943 BP; 311 A; 199 C; 199 G; 229 T; 5 other:

Alignment_scores:

Quality: 194.00 Length: 194
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:
US-09-674-266A-181 x AAH35026 ..

Align seg 1/1 to: AAH35026 from: 1 to: 943

13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
|||||
26 CCGACTCTAGCCGCCGACTCACACAAAGCAGGTGGGTGAGAAATCCAG 75
29 gValAlaMetGluLysIleProValSerLAsPheLeuLeuLeuValAlaL 46
|||||
76 AGTTGCCATGGAGAAATTCACAGTGCACATCTTGCTCCCTGGGCC 125
46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
|||||
126 TCTCTACACTCTGGCCAGATACACAGTCACAACTGAGACCAAAAG 175
63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
|||||
176 GACACAAAGACTCTGCAGCCAACTGCCAGACCTCTCCAGAGGTG 225
79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGlnGluAlaLeuTyrLys 96
|||||
226 GGTGACCAACTCATCTGAGCTCAGACATATGAGAAAGCTTATATAAT 275
96 eLysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGlyLys 112
|||||
276 CCAAGACAAAGCAACAACCTTGATGATTATTCATCAGTGGATGAGTGC 325
113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleG 129
|||||
326 CCACACAGTCAAGCTTTAAAGAAAGTGTGCTGAAATAAAGAAATCCA 375
129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluTrpThr 146
|||||
376 GAAATGGCAGAGAGTGTGCTCCTCAATGCTTATGTAACAAACG 425
146 sPLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
|||||
426 ACAACACCTTCTCCTGATGGCCAGTATGTCGCCAGATTTGTTGTT 475
163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
|||||
476 GACCATCTCTGACAGTTAGAGCCGATATCAGTGAAGATATTCAAAYG 525
179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMet 196
|||||
526 TCTGTAAGTTAGAACCTGCAGATACAGCTGCTGTTGCTTGACAAATGA 575
196 ySLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
576 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 607
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ77486
seq_documentation_block:
ID AAZ77486 standard; cDNA: 1018 BP.
XX
XX AAZ77486:
XX
XX 10-APR-2000 (first entry)
XX
XX Human ovarian tumor cDNA library derived EST fragment 37.
DE Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KM gene therapy; treatment; ss.
XX
XX Homo sapiens.
OS
XX DE19817557-A1.
PN
XX 21-OCT-1999.
PD
XX 09-APR-1998; 98DE-1017557.
XX
XX 09-APR-1998; 98DE-1017557.
PF
XX 09-APR-1998; 98DE-1017557.
PR
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
PA

CC colon cancer. huxAG-1 cDNA was isolated from a cDNA library
 CC derived from human colon cancer tissue. Vectors, host cells,
 CC antibodies, and screening methods for identifying agonists and
 CC antagonists of huxAG-1 are provided. HuxAG polypeptides are growth
 CC factors and can be used to stimulate proliferation of cells. They
 CC can be used to stimulate the proliferation and differentiation of
 CC hepatocytes to alleviate or treat liver diseases and pathologies
 CC caused by fulminant liver failure caused by cirrhosis, liver damage
 CC caused by viral hepatitis and toxic substances. They can also be
 CC used to stimulate or promote liver regeneration, e.g. after
 CC surgery. They can also be used to prevent and heal damage to the
 CC lungs caused by various pathological states. They can be used to
 CC stimulate proliferation and differentiation and promote the repair
 CC of alveoli and bronchiolar epithelium to prevent, attenuate, or
 CC treat acute or chronic lung damage, e.g. emphysema, which results
 CC in the progressive loss of alveoli, and inhalation injuries, e.g.
 CC resulting from smoke inhalation and burns, that cause necrosis of
 CC the bronchiolar epithelium and alveoli. They can also be used to
 CC stimulate the proliferation and differentiation of breast tissue
 CC and could therefore be used to promote healing of breast tissue
 CC injury due to surgery, trauma or cancer. Antagonists can be used
 CC to treat hyperproliferative disorders, including cancer, in
 CC particular hepatocellular carcinoma, osteoclastoma, breast cancer,
 CC or colon cancer. The products can also be used for detection and
 CC diagnosis.

XX Sequence 875 BP; 283 A; 189 C; 180 G; 223 T; 0 other;

alignment_scores:

Quality: 192.00 Length: 192
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266A-181 x AAV19155 ..

Align seg 1/1 to: AAV19155 from: 1 to: 875

15 ProSerArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
 |||||||
 20 CCGAGCCGCCACATCAGCAGGAGGCGTGAGAAATCCAGAGTTGC 69
 |||||||
 31 aMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuSert 48
 |||||||
 70 CATGAGAGAAATTCACGATTCCTCTCTGTCCTCTCTCTCTCTCT 119
 |||||||
 48 yTrpHleAlaArgAspThrValLysProGlyAlaLysLysAspThr 64
 |||||||
 120 ACACCTGCGCAGAGATACACAGTCAAACTGAGCCAAAAGGACACA 169
 |||||||
 65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
 |||||||
 170 AAGGACTTCACCCAAATCCGCCAGACCTCTCCAGAGGTGGGGTGA 219
 |||||||
 81 pGlnLeuIleTrpThrGlnThrTyrgLugluAlaLeuTyrlsSerLysT 98
 |||||||
 220 CCAACTCATCTGGACTCAGACATATGAGAAGCTTATATTAATCCAGA 269
 |||||||
 98 hrSerAsnLysProLeuMetIleIleHisIleLeuAspGluCysProHis 114
 |||||||
 270 CAAAGCAAAACCTTGATGATTAATTCATCTGAGATGAGCGCCACAC 319
 |||||||
 115 SerGlnAlaLeuLysLysValPheAlaGlnLysLysGluIleGlnLysLe 131
 |||||||
 320 AGTCAAGCTTTAAAGAAAGTGTGCTGAATAAATAAGAAATCCAGAAAT 369
 |||||||
 131 uAlaGlnLysPheValLeuLeuAsnLeuValTyrgLutThrAspLysH 148
 |||||||
 370 GCGAGAGAGATTGCTCTCAATCTGTTATGAACAACACTGACAAAC 419
 |||||||
 148 lsLeuSerProAspGlyLysIleTyrgValProArgIleMetPheValAspPro 164
 |||||||

420 ACCTTCTCCTGATGCGCAGATATGTCCCGAGATTATGTTGTGACCCA 469
 165 SerLeuThrValAlaArgAlaAspIleThrGlyArgTySerAsnArgLeuTy 181
 |||||||
 470 TCTCTGACAGTTAGAGCCGATATCATCGAAGATATTAATCTCTCTCTA 519
 |||||||
 181 rAlTyrgLysProAlaAspThrAlaLeuLeuAspAsnMetLysLysA 198
 |||||||
 520 TGCTTACGAACTCGCAGATACGCTGTGCTTGACAAACATGAAGAAG 569
 198 lAlaLysLeuLeuLysThrGluLeu 206
 |||||||
 570 CTCCTCAAGTCTCTCAAGACTCAATTG 595
 seq_name: /sids1/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF63314

seq_documentation_block:

ID AAF63314 standard; cDNA; 875 BP.

AC AAF63314;

DT 04-MAY-2001 (first entry)

DE Human huxAG-1/CCSG colon cancer specific gene cDNA.

XX Human; growth factor; huxAG-1; colon cancer specific gene; CCSG;

KW cell proliferation; liver disease; fulminant liver failure; cirrhosis;

KW hepatitis; cancer; colon cancer; colorectal carcinoma; ss.

XX Homo sapiens.

PN US6171816-B1.

XX 09-JAN-2001.

PD 22-AUG-1997; 97US-0916576.

XX 23-AUG-1996; 96US-0024347.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Yu G, Dillon PJ, Ebner R, Endress GA;

DR WPI; 2001-136872/14.

DR P-PDSB; AAB72203.

PT Novel human growth factor polypeptide useful for diagnosing and

XX treating colon cancer and liver diseases, to prevent and heal damage to

XX the lungs and for identifying modulators of therapeutic use -

XX Claim 45; Fig 1; 59pp; English.

XX This invention relates to a human growth factor polypeptide huxAG-1 also

XX known as a colon cancer specific gene (CCSG). HuxAG-1 stimulates cell

XX proliferation as a growth factor. The HuxAG-1 protein is useful for

XX identifying compounds capable of enhancing or inhibiting cellular

XX response induced by huxAG-1. The protein is also useful for stimulating

XX proliferation of cells e.g. colon, breast, liver and lung cells, and

XX hepatocytes. It is useful for alleviating or treating liver diseases and

XX pathologies such as fulminant liver failure caused by cirrhosis, liver

XX damage caused by viral hepatitis and toxic substances, for preventing and

XX treating damage to the lungs caused by various pathological states and

XX for promoting healing of breast tissue injury due to surgery, trauma or

XX cancer. HuxAG-1 and the identified antagonist are useful for treating

XX cancer, in particular colon cancer. Detecting altered levels of huxAG-1

XX and its polynucleotides are useful for diagnosing or detecting cancer in

XX mammals. The gene encoding huxAG-1 is useful for monitoring human

XX colorectal carcinoma. huxAG-1 nucleic acid molecules are also useful for

XX chromosome identification. The present sequence represents cDNA encoding

XX huxAG-1.

XX Sequence 875 BP; 283 A; 189 C; 180 G; 223 T; 0 other;

alignment_scores:

Quality: 192.00 Length: 192
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x AAF63314 ..

Align seg 1/1 to: AAF63314 from: 1 to: 875

```

15 ProSerArgLeuThrGlnGlyArgTyrValArgLysSerArgValAla 31
   |||||
20 CCRAGCCGCCGACTCACACAGCAGGTGGGTGAGAAAACCGAGTGGC 69
   |||||
31 aMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerT 48
   |||||
48 yTrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
   |||||
120 ACACCTGTGGCCAGAGATACCACAGTCAAACTGGACCCAAAAGGACACA 169
   |||||
65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTyrPglYas 81
   |||||
170 AAGGACTCTCGACCCAACTGCCCCAGACCCCTCTCCAGAGTGGGGTGA 219
   |||||
81 pGlnLeuIleTyrPThrGlnThrTyrgLugLualLeuTyrLysSerLysT 98
   |||||
220 CCNAACATCTGAGCTCAGACATATGAGAAAGCTTATATATAATCCAGA 269
   |||||
98 hSerSerLysProLeuMetIleIleHisIleLeuAspGluCysProHis 114
   |||||
270 CAAGCAACAAACCTTGATGATTAATTCATCTGATGATGAGTCCCAAC 319
   |||||
115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLys 131
   |||||
320 AGTCAGGCTTTAAGAAAGTGTGCTGAAATAAAGAAATCCAGAAAT 369
   |||||
131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
   |||||
370 GGCAGAGCAGATTGTCTCTCTCAATCTGTGTATGAACAACAGACAAAC 419
   |||||
148 IsteuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
   |||||
420 ACCCTTCTCTGATGGCAGTATGTCCCGAGATTAATGTTGTGACCCA 469
   |||||
165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
   |||||
470 TCTCTGACAGTTAGAGCCGATATCAGTGGAGATATTCAATGCTCTCTA 519
   |||||
181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
   |||||
520 TGGTTTACGAACCTGCAGATACAGCTCTGTGTGTCACACATGAGAGAA 569
   |||||
198 lalaLeuLysLeuLeuLysThrGluLeu 206
   |||||
570 CTCTCAAGTTGCTGAAGACTGAATTTG 595

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV59320

seq_documentation_block:

ID AAV59320 standard: cDNA, 881 BP.

XX AAV59320:

DT 21-DEC-1998 (first entry)

DE Nucleotide sequence encoding zsig10 polypeptide.

XX ss: human; mucous-mediated function; adhesion; tumour metastasis;
 KW bacterial colonisation; microbial infection; AIDS; cystic fibrosis;
 KW chronic obstructive pulmonary disease; asthma; Crohn's disease;

KW sinonasal inflammatory disease; inflammatory bowel disease; bronchitis.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 63..590

FT /tag= a

FT /product= "zsig10"

XX W09841627-A1.

XX PD 24-SEP-1998.

XX PF 18-MAR-1998; 98WO-US05251.

XX PR 19-MAR-1997; 97US-0039631.

XX PA (ZYMO) ZYMOGENETICS INC.

XX PI Sheppard PO;

XX DR WPI; 1998-531566/45.

XX DR P-PSDB; AAW77365.

PT New isolated mucous-associated polypeptide, zsig10 - used to develop products for treating e.g. tumour metastasis, microbial infections, cystic fibrosis, asthma, bronchitis or inflammatory bowel disease

PS Claim 37; Page 80-81; 109pp; English.

CC The human polypeptide zsig10 is involved in mucous-mediated functions such as adhesion. The products of the invention can be used in the study and treatment of e.g. tumour metastasis, bacterial colonisation, susceptibility to and persistence of infection, microbial infections, CC AIDS, cystic fibrosis, chronic obstructive pulmonary disease, asthma, CC sinonasal inflammatory disease, inflammatory bowel disease, bronchitis, CC or Crohn's disease. The products can also be used for detection, diagnosis and drug screening.

XX SQ Sequence 881 BP; 293 A; 185 C; 180 G; 223 T; 0 other;

alignment_scores:

Quality: 191.00 Length: 191
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x AAV59320 ..

Align seg 1/1 to: AAV59320 from: 1 to: 881

```

16 SerArgArgLeuThrGlnGlyArgTyrValArgLysSerArgValAla 32
   |||||
15 AGCCGCCGACTCACACAGCAGGTGGGTGAGAAAACCGAGTGGCAT 64
   |||||
32 tGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerTyrT 49
   |||||
65 GGAGAAAATTCAGTGCAGATTCCTGCTCTGTGGCCCTCTCTTACA 114
   |||||
49 hLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThrLys 65
   |||||
115 CTCTGGCCAGATACCAAGTCCAACTGGACCCAAAAGGACACAAAG 164
   |||||
165 GACTCTGACCCAAACTGCCAGACCCCTCTCCAGAGTGGGGTGACCA 214
   |||||
66 AspSerArgProLysLeuProGlnThrLeuSerArgGlyTyrPglYas 82
   |||||
82 nLeuIleTyrPThrGlnThrTyrgLugLualLeuTyrLysSerLysThr 99
   |||||
215 ACTCATCTGAGCTCAGCATATGAGAAAGCTTATATAATAATCCAGACAA 264
   |||||
99 eAsnLysProLeuMetIleIleHisIleLeuAspGluCysProHisSer 115

```

```
|||||
265 GCAACAAACCTTGATGATTTATTCATCCTGGATGAGTCCACACAGT 314
116 GlnAlaLeuLysValPheAlaGluAsnLysGluIleGlnLysLeuAl 132
315 CAACCTTTAAAGAAAGTGTTCGTAAGAAATTAAGAAATCCAGAAATTGGC 364
132 aGluGlnPheValLeuLeuAsnLeuValTYrGluThrThraspLysHSL 149
365 AGACACAGTTTGCTCCTCCATCTGTTTATGAAGAAACAACCTACCAACGCC 414
149 euserProaspGlyGlnTYrValProArGileMetPheValaspProser 165
415 TTTCTCTGATGGCCAGTATGTCCCCAGATTAATGTTTGTGACCCATCT 464
166 LeuThrValAlaArgAlaAspIleThrGlyArgTYrSerAsnArgLeuTYrAl 182
465 CTGACAGTTNAGCCGATATCAGTCGAAAGATATTCAAATCCTCTATAGC 514
182 aTYrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaL 199
515 TTACGAACCTGCAGATACAGCTCTGTTGCTTGACACATGAGAAAGCTC 564
199 eLysLeuLeuLysThrGluLeu 206
565 TCAAGTTGCTGAAGACTGAATTG 587
```

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: AAS13480

seq_documentation_block:

ID AAS13480 standard; DNA: 543 BP.

```
XX AAS13480:
AC
XX
DT 19-DEC-2001 (first entry)
XX
DE DNA encoding breast cancer cell membrane protein 7 (BCMP 7).
XX
KM Breast cancer cell membrane protein 7; BCMP 7; breast cancer;
KW metastasis; cyclostatic.; gene therapy; antibody; antisense;
KM vaccine; chromosome 7p21.3; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 11..538
FT FT /*tag- a
FT FT /product= "BCMP 7"
FT FT /note= "Breast cancer cell membrane protein 7"
FT FT sig_peptide 11..70
FT FT /*tag- b
FT FT mat_peptide 71..535
FT FT /*tag- c
FT FT /note= "Mature BCMP 7"
XX
PN WO200163290-A1.
XX
XX 30-AUG-2001.
XX
XX 21-FEB-2001; 2001WO-GB00734.
XX
XX 25-FEB-2000; 2000GB-0004576.
XX
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
XX PA Boyd RS, Stamps AC, Terrett JA, Tyson KL;
XX PI WPI; 2001-570651/64.
XX DR P-PSDB; AAU08804.
XX
XX Diagnosing, preventing and treating breast cancer using a breast cancer
XX cell membrane protein BCMP 7 -
```

PS Claim 6; Fig 1; 62pp; English.

XX The invention describes the novel use of a protein found in breast
CC cancer cell membranes (BCMP 7) for diagnosing, preventing and treating
CC breast cancers. The peptide has cyclostatic action and potential uses in
CC gene therapy and in vaccines. The polypeptide, antisense nucleic acids,
CC or fusion proteins comprising and green fluorescent protein or the DsRed
CC fluorescent protein, antibodies specific for and/or nucleic acid are
CC used for the prevention and/or treatment of breast cancer. Antibodies
CC against may also be used for screening for and/or diagnosis of breast
CC cancer in a patient. The method for monitoring/assessing breast cancer
CC treatment in a patient and for the identification of metastatic breast
CC cancer cells in samples from a patient. This sequence encodes breast
CC cancer cell associated protein 7 (BCMP 7), located on chromosome
CC 7p21.3, described in the method of the invention.

XX Sequence 543 BP; 170 A; 135 C; 107 G; 131 T; 0 other;

alignment_scores:
Quality: 178.00 Length: 178
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266A-181 x AAS13480 ..

Align seg 1/1 to: AAS13480 from: 1 to: 543

```
29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
|||||
2 AGAGTTGCCATGAGAGAAATCCAGTGTACAGATTCTTGCTCTTGCGC 51
45 aLeuSerTYrThrLeuAlaArgAspThrValLysProGlyAlaLysL 62
|||||
52 COTCTCCTACACTGTGGCCAGATACACAGTCAAACTGAGCCAAAA 101
62 yAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
|||||
102 AGCAGACAAAGAGACTGTGACCCAAACGCCCCACACCTCTCCAGAGT 151
79 TyrGlyAspGlnLeuIleTyrThrGlnThrTYrGluGluAlaLeuTYr 95
|||||
152 TGGGTGACCACTCATCTGAGCTCAGACATRTGAAGAACTCTATATA 201
95 sSerLysThrSerAsnLysProLeuMetIleIleHisLysLeuSpGluC 112
|||||
202 ATCCAGACAAAGCAACAAACCTTGATGATTTATTCATCCTGGATGAGT 251
112 ySProHisSerGlnAlaLeuLysValPheAlaGluAsnLysGluLle 128
|||||
252 GCCCAGACAGTCACAGCTTTAAAGAAAGTGTTCGTAAGAAATTAAGAAATTC 301
129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTYrGluThrTh 145
|||||
302 CAGAAATTTGGCAGACAGTGTGCTCCTCCTCAATCTGTTTATGAAGAAC 351
145 rAspLysHisLeuSerProaspGlyGlnTYrValProArGileMetPheV 162
|||||
352 TGACAAACACCTTCTCCTGATGGCCAGATGTCCCCAGAGATATGTTG 401
162 aLaspProSerLeuThrValArgAlaAspIleThrGlyArgTYrSerAsn 178
|||||
402 TTGACCCATCTCTGACAGTTAGAGCCGATATCAGTCGAAAGATATTCAAAC 451
179 ArgLeuTYrAlaTYrGluProAlaAspThrAlaLeuLeuAspAsnMe 195
|||||
452 GCTCTCTATGCTTACGAACCTGACAGATACAGCTCTGTTGCTTGACAAAT 501
195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
|||||
502 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 535
```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ40803

seq_documentation_block:

ID AAZ40803 standard; DNA: 1689 BP.

AC AAZ40803;

DT 18-JAN-2000 (first entry)

DE Secreted protein EST coding sequence 108-008-5-0-A6-FL.

XX Secreted protein; fingerprint identification technique;

KM chromosome mapping; human; hereditary disease; diagnosis;

KM hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;

KM autoimmune disease; rheumatic disease; embryonic disorder; myopathy;

KM renal injury; amino aciduria; hypoglycaemia; male rat infertility;

KM hypertension; ss.

XX Homo sapiens.

OS Homo sapiens.

PN W09940189-A2.

XX 12-AUG-1999.

PD 09-FEB-1999; 99MO-1B00282.

XX 09-FEB-1999; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

PR 10-AUG-1998; 98US-0096116.

PR 04-SEP-1998; 98US-0099273.

XX (GEST) GENSET.

XX Bouquelier L, Duclet A, Dumas Milne Edwards J;

DR WPI, 1999-600966/51.

XX P-PSDB; AAV59675.

PT Extended cDNAs useful for expressing secreted proteins and to obtain

XX specific antibodies -

PS Claim 1: Page 168-169; 244p; English.

XX This sequence encodes a human secreted protein of the invention. The
CC extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC prepare PCR primers and probes. These are useful for forensic matching or
CC positive identification by DNA sequencing. They may also be used in
CC alternative fingerprint identification techniques. Antibodies against the
CC proteins encoded by the extended cDNAs are useful in identification of
CC tissue types or cell species, as well as identifying tissue specific
CC soluble proteins. The sequences can be used for chromosome mapping and
CC identification of genes associated with hereditary diseases or drug
CC response. Signal sequences from the cDNAs can be used in construction of
CC secretion vectors. Other sequences derived from the extended cDNAs can be
CC used to clone upstream genomic DNA sequences including promoters. This is
CC in turn useful for identifying proteins that interact with promoter
CC sequences. Some of the proteins may be useful in diagnosing and treating
CC several disorders including, but not limited to: cancer, hyperlipidaemia,
CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC rheumatic diseases, embryonic disorders, hypertension, renal injury,
CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.

XX Sequence 1689 BP; 552 A; 350 C; 335 G; 452 T; 0 other;

alignment_scores:

Quality: 178.00

Ratio: 1.000

Percent Similarity: 100.000

Length: 178

Gaps: 0

Percent Identity: 100.000

alignment_block:

US-09-674-266A-181 x AAZ40803

Align seq 1/1 to: AAZ40803 from: 1 to: 1689

29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
42 AGAGTGGCCATGAGAAATTCAGAGTGCAGCATCTTCTGCTGTCG 91
45 aleuSerTyrThrLeuAlaArgAspThrValLysProGlyValLysL 62
92 CCTCTCTACACTGTGGCAGAGATACCAAGTCAAACTGGAGCCAAA 141
62 ysaAspThrLysAspSerArgProLysLeuProGlnThrLysSerArg 78
142 AGGACAAAGAAAGACTCTCGAACCAACTGCCAGACCTCTCAGAGT 191
79 TrpGlyAspGlnLeuIleTyrPheGlnThrTyrGluGluAlaLeuTyr 95
192 TGGGGTGCACCACTCATCTGACTCAGACATATGAAACACTCATAT 241
95 sSerLysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGlu 112
242 ATCCAGACAAAGCAACAAACCTTGATGATATTCATCACTGATGAGT 291
112 yspProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGlu 128
292 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTCGTGAATAAAGAAATC 341
129 GlnLysLeuAlaGluGlnPheValLeuAsnLeuValTyrGluThrTh 145
342 CAGAAATTTGACAGAGCAGTTGTCTCTCAATCTGTTATGAAACAC 391
145 rasPlyHisLysSerProAspGlyGlnTyrValProArgIleMetPhe 162
392 TGACAAACACCTTTCTCTGATGAGCAGATGTCCTCCAGGATTAAT 441
162 alaSpProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
442 TTGACCATCTCTGACAGTTAGAGCCGATATCACTGAGATATTCAAT 491
179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuAspAsnMe 195
492 CGTCTCTATGCTTACGAACTGCAGATACAGCTCTGTGCTTGACAA 541
195 tlyLysAlaLeuLysLeuLysThrGluLeu 206
542 GAGAAAGCTCTCAAGTTGCTGAGAGACTGAATTG 575
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV29048
seq_documentation_block:
ID AAV29048 standard; cDNA: 525 BP.
XX AAV29048;
AC AAV29048;
DT 21-AUG-1998 (first entry)
XX Open reading frame human protein comprising secretory signal 9.
DE Human protein; secretory signal; nutritional source; cytokine;
KW immunity; hemotopiasis; activin; inhibin; tumour; chemotactic;
KW chemokinetic; thrombolytic; anti-inflammatory; inhibition;
KW stomach cancer cell; ds.
XX Homo sapiens.
OS Homo sapiens.
XX Key Location/Qualifiers
FH 1..525
FT CDS /tag= a
FT /product= "human protein comprising secretory
FT signal"
XX W09811217-A2.
PN 19-MAR-1998.

XX 12-SEP-1997; 97WO-JP03239.
 XX
 XX 13-SEP-1996; 96JP-0243060.
 XX
 PA (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 XX
 XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;
 PI
 DR WPI: 1998-207380/18.
 DR P-PSDB; AAM37872.

PT Human proteins with secretory signal sequences - used to treat
 PT Immune deficiencies, infections, tumours, and haematopoietic
 PT disorders, etc.

PS Claim 3; Pages 88; 131pp; English.

XX This is the nucleotide sequence of the open reading frame of a novel
 CC human protein comprising a secretory signal (AAV29047), isolated from
 CC stomach cancer cells. Its proteins can be used as nutritional sources
 CC or supplements. The proteins may also have cytokine functions,
 CC immune modulating functions, haematopoiesis regulating activity,
 CC activin/inhibin regulating activity, chemotactic/chemokinetic activity,
 CC haemostatic and thrombolytic activity, receptor/ligand activity,
 CC anti-inflammatory activity, tumour inhibition activity.

XX Sequence 525 BP; 162 A; 131 C; 103 G; 129 T; 0 other;

alignment_scores:
 Quality: 175.00 Length: 175
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x AAV29048

Align seg 1/1 to: AAV29048 from: 1 to: 525

32 MetGluIySIIeProValSerAlaPheLeuLeuValAlaLeuSerTy 48
 1 AAGGAGAAATTCAGTGCAGATTCCTGCTGTCCTGCTCCTA 50
 48 rThreuaLaArGAsPThrThrValysProGlyAlaLysLysAspThrL 65
 51 CACTCTGGCCAGATACACAGTCAAACTGGAGCCAAAAGGACACAA 100
 65 ysAspSerArGProLysLeuProGlnThrLeuSerArGlyTrpGlyasp 81
 101 AGGACTCTGCAGCCAAACTGCCAGACCCCTCTCCAGAGTGGGGTGAC 150
 82 GlnLeuIleTrpThrGlnThrTyrrGluGluAlaLeuTyrrLysSerLysH 98
 151 CAACGCACTCTGGACTCAGACATATGAGAAAGCTCTATATAAATCCAAAGAC 200
 98 rSerAsnLysProLeuMetIleIleHisLysLeuAspGluCysProHisS 115
 201 AAGCAACAACCCCTGATGATATTCATCACCCTTGGATGAGTCCACACA 250
 115 erGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeu 131
 251 GTCAAGCTTTAAAGAAAGTGTGCTGAATAATAAAGAAATCCAGAAATTG 300
 132 AlaGluGlnPheValLeuLeuAsnLeuValTyrrGluThrThraspLysH 148
 301 GCAGAGCAGTTTGTCTCTCAATCTGGTTATGGAACCAACGACAAACA 350
 148 sLeuSerProAspGlyGlnTyrrValProArgIleMetPheValasProS 165
 351 CTTTCTCTGATGGCCAGATATGCCAGAGATTATGTTGTGACCCAT 400

165 erLeuThrValaLarAlaAspIleThrGlyArgTySerAsnArgLeuTy 181
 401 CTCTGACAGCTAGAGCCGATATCAGTGAAGATATTCAAACCGTCTCAT 450
 182 AlaTyrrGluProAlaAspThrAlaLeuLeuAsnMetLysLysAl 198
 451 GCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACACATGAAAGAAC 500
 198 aLeuLysLeuLeuLysThrGluLeu 206
 501 TCTCAAGTGTCTGAGACATGATTG 525

seq_name: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT.AAC00115

seq_documentation_block:

ID AAC00115 standard; cDNA; 468 BP.

XX AAC00115;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 113.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX P-PSDB; AAG00109.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 113; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

XX cDNAs encoding secreted proteins. An ORF has been identified within the

XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

XX derived from 30 different tissues. EST sequences usually correspond

XX mainly to the 3' untranslated region (UTR) of the mRNA because they are

XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

XX well suited for isolating cDNA sequences derived from the 5' ends of

XX cDNAs and even in those cases where longer cDNA sequences have been

XX obtained, the full 5' UTR is rarely included. 5' ESTs are derived from

XX cDNAs with intact 5' ends and can therefore be used to obtain full length

XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,

XX gene therapy and chromosome mapping procedures. They are used to obtain

XX upstream regulatory sequences and to design expression and secretion

XX Sequence 468 BP; 142 A; 122 C; 98 G; 106 T; 0 other;

alignment_scores:
 Quality: 146.00 Length: 146
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x AAC00115 ..

Align seg 1/1 to: AAC00115 from: 1 to: 468

```

12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
13 GGACCCGATCGTAGCCGCGACTCACACAAAGCAGGTGGTGAAGAAATC 62
28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValA 45
63 CAGAGTTGGCATGGAGAAATTCAGCTGCAGCATTCCTGCTCTGTGG 112
45 ILeuSerTrpThrLeuAlaArgAspThrThrValLysProGluAlaLys 61
113 CCTCTCTACACTCTGGCCAGAGATCCACAGTCAAACTGAGCCAA 162
62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgI 78
163 AAGGACACAAAGGACTCTGCAGCCAAAGTCCAGACCCCTCCAGAG 212
78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrgluGluAlaLeuTyrL 95
213 TTGGGGTGACCAACTATCTGGACTGACATGTGAAGAGCTCTATATA 262
95 ySerSerLysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGlu 111
263 AATCCAGACAGCAACAAACCTTGATGATTTATTCATCTGATGATGAG 312
112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluI 128
313 TGCCACACAGTCAGCTTAAAGAAAGTGTGTGCAAAATGAAGAAAT 362
128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
363 CCGAAATGGCGAGCAGAGTTGTCTCTCAATCTGGTTATGAACAA 412
145 hrAspLysHisLeuSerProAspGlyGlnTyrValPro 157
413 CTGACAAACACCTTCTCTCGATGGCCAGATGTCCCA 450

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF44884

seq_documentation_block:

ID AAF44884 standard; cDNA; 778 BP.

AC AAF44884;

DT 28-MAR-2001 (first entry)

DE Human breast cancer related protein coding sequence SEQ ID NO: 40.

KW Human; breast cancer; diagnosis; therapy; vaccine; ss.

OS Homo sapiens.

PN WO200078960-A2.

PD 28-DEC-2000.

PF 23-JUN-2000; 2000WO-US17536.

PR 23-JUN-1999; 98US-0140903.

PR 12-OCT-1999; 99US-0138980.

XX (CORI-) CORIXA CORP.

PI Yuqiu J, Mitcham JL;

DR WPI; 2001-041426/05.

PT New polynucleotides encoding breast tumor specific proteins, useful for prevention, treatment and diagnosis of breast cancer -

XX

PS Claim 25; Page 132; 165pp; English.

XX

CC The present invention provides the coding sequences for a number of breast cancer related proteins. These can be used in vaccinations against, diagnosis of and treatment of cancer, particularly breast cancer.

XX

SO Sequence 778 BP; 245 A; 175 C; 158 G; 193 T; 7 other;

alignment_scores:

Quality: 134.00 Length: 134

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x AAF44884 ..

Align seg 1/1 to: AAF44884 from: 1 to: 778

```

54 ThrThrValLysProGluAlaLysLysAspThrLysAspSerArgProly 70
125 ACCACAGTCAAAACCTGAGCCAAAGGACACAAAGAGACTCTGACCCAA 174
70 sleuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIleTrpThrG 87
175 ACTGCCCCAGACCCCTCTCCAGAGGTTGGGTGACCAACTATCTGAGACT 224
87 ILeuTrpGluGluAlaLeuTyrLysSerLysThrSerAsnLysProLeu 103
225 AGACATATGAAAGAGCTCTATATAATCCAAAGCAAGCAACAAACCTTG 274
104 MetIleIleHisLysLeuAspGluCysProHisSerGlnAlaLeuLys 120
275 ATATATTATTCATCACTGGATGAGTGGCCACACAGTCAACCTTTAAAGAA 324
120 sValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGlnPheVal 137
325 AGTGTGTTGCTGAATAATTAAGAAATCCAGAAATGGCAGACAGCTTGTC 374
137 eLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerProAspGly 153
375 TCCTCATCTGGTTTATGAACAACAGTACAAACACCTTCTCTCGATGGC 424
154 GlnTyrValProArgIleMetPheValAspProSerLeuThrValArgI 170
425 CAGTATGTCGCCAGAGATTATGTTGGACCACTCTGTGACAGTTAGAGC 474
170 aAspIleThrGlyArgTrpSerAsnArgLeuTyrAlaTyrGluProAla 187
475 CGATATCACTGAGAGATATTCAATGCTCTATGCTTACGAACCTGCAG 524
187 sp 187
II
525 AT 526

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA77903

seq_documentation_block:

ID AAA77903 standard; cDNA; 409 BP.

AC AAA77903;

DT 14-NOV-2000 (first entry)

DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:183.

KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine; immunotherapy; diagnosis; progression; ss.

OS Homo sapiens.

PN WO200037643-A2.

XX

```

XX 29-JUN-2000.
PD
XX
XX 23-DEC-1999: 99WO-US30909.
PF
XX 23-DEC-1998: 98US-0221298.
PR 02-JUL-1999: 99US-0347496.
PR 22-SEP-1999: 99US-0401064.
PR 19-NOV-1999: 99US-0444242.
PR 02-DEC-1999: 99US-0454150.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk J;
PI Wang T, Yugu J;
XX
XX WPI: 2000-442671/38.
XX
XX New colon tumor polypeptides used to inhibit the development of cancer,
PT especially colon cancer, and for diagnosing and monitoring the
PT progression of the cancer -
XX
XX
XX Claim 29: Page 144: 229pp: English.
XX
XX Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
CC portions of proteins which are associated with human colon tumours.
CC The invention also specifically discloses 8 human colon tumour proteins
CC (AA81897-811904). The nucleic acids, the polypeptides they encode, and
CC antigen presenting cells (APCs, preferably dendritic cells) expressing
CC such polypeptides may be used in vaccines that target tumour cells,
CC especially colon tumour cells, thereby inhibiting the development of
CC cancer. T-cells specific for the polypeptide expressed by the APC are
CC used to remove tumour cells from biological samples, especially blood or
CC fractions thereof. The sample or the isolated T-cells specific for the
CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
CC CD8+ T-cells from a patient may be incubated with a polypeptide or
CC nucleic acid of the invention, or an APC expressing such a polypeptide,
CC to cause the proliferation of specific T-cells. The T-cells can be
CC cloned and then administered back to the patient to inhibit cancer
CC development. Nucleic acids encoding the polypeptides and antibodies
CC against the polypeptides may be used to determine the expression level
CC of a tumour protein of the invention, and therefore to determine whether
CC cancer cells are present. Such diagnostic methods may also be used to
CC monitor the progression of a cancer by repeating the processes at time
CC intervals, and comparing the current result to previous results. The
CC present sequence represents a cDNA encoding a human colon tumour
CC polypeptide.
XX
XX Sequence 409 BP: 131 A; 104 C; 80 G; 94 T; 0 other;

```

```

alignment_scores:
  Quality: 133.00      Length: 133
  Ratio: 1.000         Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment_block:

US-09-674-266A-181 x AAA77903 ..

Align seg 1/1 to: AAA77903 from: 1 to: 409

```

48 TyrThrLeuAlaArgAspThrThrVallyProGlyAlaLysLysAspTh 64
|||||
11 TACACTCTGGCCAGATACCACTCAAACTGAGCCAAAGAGAC 60
|||||
64 rlyAspSerArgProLysLeuProGlnThrLeuSerArgGlyTPGlyA 81
|||||
61 AAAGGAGCTCGAGCCAAAGTCCCAACCCCTCTCCAGAGGTGGG 110
|||||
81 spGlnLeuIleTrpThrGlnThrTrpGluGluAlaLeuTyrLysSer 97
|||||
111 ACCAACTATCTGGACTCAGACATATGAAGAAAGCTCTATATAATCC 160

```

```

98 ThrSerAsnLysProLeuMetIleHisHisLeuAspGluCysProH 114
|||||
161 AAGAGCAACAAACCTTGATGATTCATCATCTGATGATGCCACA 210
|||||
114 sSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGln 131
|||||
211 CAGTCAGCTTTAAAGAAAGTGTGGCTGAAATAAAGAAATCCGAAAT 260
|||||
131 euAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAsp 147
|||||
261 TGGCAGACAGATTGTCTCTCTCATATCGGTTTANGAACAACAGTACA 310
|||||
148 HisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 164
|||||
311 CACCTTCTCCGATGAGCCAGTATGTCGCCAGATTAATGTTGTGACCC 360
|||||
164 oSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeu 180
|||||
361 ATCTGTGACAGTTAGAGCCGATATCCTGAGAGATATTCATTCGCTTC 409
|||||
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA128641
seq_documentation_block:
ID AA128641 standard: cDNA: 409 BP.
XX
XX AA128641:
XX
XX 12-OCT-2001 (first entry)
XX
DE Colon tumour related determined cDNA sequence for C1140.
XX
XX Human: immunotherapy; diagnosis: colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colonic cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
PF 29-DEC-2000: 2000WO-US35596.
XX
PR 30-DEC-1999: 99US-0476296.
PR 10-JAN-2000: 2000US-0480321.
PR 15-FEB-2000: 2000US-0504629.
PR 06-MAR-2000: 2000US-0519444.
PR 19-MAY-2000: 2000US-0575251.
PR 29-JUN-2000: 2000US-0609448.
PR 28-AUG-2000: 2000US-0649811.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
XX WPI: 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
PT
XX
XX Claim 25: Page 184: 472pp: English.
XX
XX The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host

```

CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
CC and AA124494 to AA124523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX

Sequence 409 BP; 131 A; 104 C; 80 G; 94 T; 0 other:

alignment_scores:

Quality: 133.00 Length: 133
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x AA128641 ..

Align seg 1/1 to: AA128641 from: 1 to: 409

48 TyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspTh 64
|||||
11 TACACTCTGCGCCAGATACACAGTCACACCTGAGCCAAAAGGACAC 60
64 rLysAspSerArgProLysLeuProGlnThrLysSerArgGlyTrpGly 81
|||||
61 AAAGGACTCTCGACCCAAACTGCCCAAGACCTCTCCAGAGTTGGGGTG 110
81 spGlnLeuIleTPTnGlnThrTyrGlnGluAlaLeuTyrLysSerLys 97
|||||
111 ACCAAGCTCATCTGACCTGACACATATGAAGAACTCATATTAATCCAG 160
98 ThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluCysProH 114
|||||
161 ACAAGCAACAACCTGATGATTAATCATCACTGGATGAGTGCACACA 210
114 sSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysL 131
|||||
211 CAGTCAAGCTTTAAAGAAAGTGTGGCTGAAATAAGAAATCCAGAAAT 260
131 euAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLys 147
|||||
261 TGGCAGAGCGAGTTGTCTCTCTCAATCTGTTATGAAACAACTGACAAA 310
148 HisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPr 164
|||||
311 CACCTTCTCTCTGATGCGCAGTATGTCGCCAGATTAATGTTGTGACCC 360
164 oSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeu 180
|||||
361 ATCTGTGACAGTTAGACCGATATCACTGGAAGATATTAATCAATCGTCTC 409

OM of: US-09-674-266a-181 to: EST: * out_format : pfs
Date: Oct 8, 2002 11:09 AM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+ .p2n.model -DEV-xlh
-O/cg2_1/USPTO.spool/US09674266/runat_08102002_093408_26991/app_query.fasta-1.271
-DB=EST -QMT=fastap -SUFFIX=olip2n.rst -GAPOP=4.500
-GAPOP=4.500 -MISMATCH=0.100 -LOOP=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -MISMATCH=0.050 -GAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -YGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DEEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MILEN=0 -MAXLEN=200000000
-USER=US09674266.ccgml.1.2564 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-674-266a-181

Database: EST: *

Database length: 13736207

Search time (sec): 1725.590000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est1:BG778248	195.00	3425.75	1.4e-181	666	1 BG778248 602666830P1 NIH_MGC_60
gb_est1:AM582256	195.00	3425.75	1.5e-181	689	1 AM582256 604-ST0212-120100-075-
gb_est1:BI523582	194.00	3408.61	1.3e-180	624	1 BI523582 603051712P1 NIH_MGC_12
gb_est2:BG179160	191.00	3354.19	1.4e-177	750	1 BG179160 602330032P1 NIH_MGC_91
gb_est2:BG748141	190.00	3337.21	1.2e-176	689	1 BG748141 601571509P1 NIH_MGC_35
gb_est1:BM129777	189.00	3320.91	9.9e-176	581	1 BM129777 1f23a08.y1 Meilston Norm
gb_est1:AM956284	189.00	3319.72	1.2e-175	721	1 AM956284 EST368354 MAGE resseq
gb_est2:BE786856	187.00	3281.72	1.5e-173	948	1 BE786856 60158166P1 NIH_MGC_7
gb_est2:BE870718	186.00	3266.09	1.1e-172	734	1 BE870718 601448570P1 NIH_MGC_65
gb_est2:GC285246	184.00	3229.47	1.2e-170	866	1 GC285246 602409699P1 NIH_MGC_91
gb_est2:GC386151	178.00	3124.68	8.5e-165	751	1 GC386151 60245333P1 NIH_MGC_15
gb_est2:BM006356	178.00	3124.61	8.5e-165	758	1 BM006356 603615055P1 NIH_MGC_11
gb_est2:BM006510	178.00	3124.21	9.0e-165	797	1 BM006510 603615461P1 NIH_MGC_11
gb_est2:BM006784	178.00	3124.16	9.0e-165	802	1 BM006784 603615217P1 NIH_MGC_11
gb_est2:BM006554	178.00	3123.74	9.5e-165	846	1 BM006554 603615521P1 NIH_MGC_11
gb_est2:BG610804	178.00	3123.67	9.6e-165	853	1 BG610804 602612442P1 NIH_MGC_60
gb_est2:BM006458	178.00	3123.58	9.7e-165	863	1 BM006458 603615186P1 NIH_MGC_11
gb_est2:BM006669	178.00	3123.34	1.0e-164	890	1 BM006669 603615373P1 NIH_MGC_11
gb_est2:BI763228	176.00	3091.12	6.3e-163	603	1 BI763228 603049976P1 NIH_MGC_11
gb_est2:BE867131	173.00	3037.90	5.8e-160	623	1 BE867131 601442937P1 NIH_MGC_65
gb_est2:BG823844	165.00	2895.28	5.1e-152	742	1 BG823844 602729503P1 NIH_MGC_15
gb_est2:BI770077	164.00	2877.15	5.2e-151	788	1 BI770077 603053412P1 NIH_MGC_12
gb_est1:AL343472	163.00	2863.22	3.1e-150	493	1 AL343472 AL343472 LTI_NFL006.P1
gb_est2:BI770277	162.00	2841.91	4.8e-149	781	1 BI770277 603055968P1 NIH_MGC_12
gb_est1:AM369331	161.00	2826.00	3.7e-148	627	1 AM369331 QV1-BN0002-061299-039-
gb_est1:AA314225	155.00	2721.31	2.5e-142	537	1 AA314225 EST186091 Colon carc
gb_est1:AA314372	149.00	2616.81	1.6e-136	449	1 AA314372 EST186091 Colon carc
gb_est1:AA314372	149.00	2616.78	1.6e-136	450	1 AA314372 EST186091 Colon carc
gb_est2:BF207999	149.00	2611.48	3.3e-136	879	1 BF207999 60186223P1 NIH_MGC_53
gb_est1:AA314146	146.00	2564.04	1.4e-133	437	1 AA314146 EST186091 Colon carc
gb_est1:AA315629	146.00	2563.53	1.5e-133	468	1 AA315629 EST187294 Colon carc
gb_est1:AM364300	146.00	2561.07	2.1e-133	637	1 AM364300 QV3-DTC01-291299-051-
gb_est2:BG259927	142.00	2487.80	2.5e-129	890	1 BG259927 602371889P1 NIH_MGC_93
gb_est2:BF747277	140.00	2458.27	1.1e-127	430	1 BF747277 RC1-BT0254-271100-121-
gb_est2:BF513917	140.00	2457.25	1.3e-127	489	1 BF513917 UT-H-BM1-smg-F-02-0-01
gb_est1:AA316233	138.00	2422.78	1.1e-125	440	1 AA316233 EST187476 HCC cell 11n
gb_est2:BF731442	137.00	2400.84	1.8e-124	755	1 BF731442 EST187482 Colon carc
gb_est1:AA315724	133.00	2334.49	8.7e-121	441	1 AA315724 EST187482 Colon carc

gb_est1:AA316967	133.00	2334.47	8.8e-121	442	1 AA316967 EST188638 HCC cell
gb_est1:AA316115	133.00	2332.51	1.1e-120	566	1 AA316115 EST187838 Colon car
gb_est1:AA315166	128.00	2246.70	6.8e-116	415	1 AA315166 EST186963 HCC cell
gb_est1:AM161619	124.00	2172.91	8.7e-112	619	1 AM161619 x146b06.x1 NCI-CGAP
gb_est2:BG501149	123.00	2153.84	1.0e-110	740	1 BG501149 602546389P1 NIH_MGC
gb_est2:BF680021	123.00	2153.77	1.0e-110	747	1 BF680021 602154740P1 NIH_MGC
gb_est2:BF680376	123.00	2152.72	1.2e-110	853	1 BF680376 602154161P1 NIH_MGC

seq_name: gb_est2:BG778248

seq_documentation_block:

LOCUS BG778248 666 bp mRNA linear EST 15-MAY-2001

DEFINITION 602666830P1 NIH_MGC_60 Homo sapiens CDNA clone IMAGE:4806679 5',

mrna sequence.

Accession BG778248

Version BG778248.1 GI:14048565

Keywords EST.

Source human.

Organism Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Reference 1 (bases 1 to 666)

Authors NIH-MGC http://mgc.nci.nih.gov/.

Title National Institutes of Health, Mammalian Gene Collection (MGC)

Journal Unpublished (1999)

Comment Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: DCTD/DRP

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L16M1656 row: 1 column: 08

High quality sequence stop: 665.

Location/Qualifiers

1. 666

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4806679"

/clone_lib="NIH_MGC_60"

/tissue_type="adenocarcinoma"

/lab_host="TDH10B (T1 phage-resistant)"

/note="Organ: prostate; Vector: pBR322 (Clontech);

Site_1: StII (ggccgcgcgcgc); Site_2: StII (ggccatcggc

); Double-stranded cDNA was prepared from cell line RNA.

5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGGCCATTTATGCG-3' and 3' adaptor

sequence: 5'-ATCTAGAGCCGCGCGCGCATG-dt(30)-BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA). Note: this is a NIH-MGC

library."

BASE COUNT 205 a 165 c 140 g 156 t

ORIGIN

alignment_scores:

Quality: 195.00 Length: 195

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x BG778248

Align seg 1/1 to: BG778248 from: 1 to: 666

12 G1PProHtsProSeRaGArGleuThGlnGlyAArgTrpAlaArglySe 28

|||||

12 GACCGCATCTACCGCCGCGCATGAGGCGGCGGCGGAGGAAATC 61

```

28  rargValaIaMetGluLysIleProValSerAlaPheLeuLeuValA 45
    |||||||
62  CAGAGTTCCATGGAGAAATTCAGTGTACGATCTTGCTCTGG 111
    |||||||
45  lalSerTyrrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
    |||||||
112  CCCCTCTCTACCTCTGCGCCAGAGTACACAGTCAACCTGGAGCCAA 161
    |||||||
62  LysAspThrLysAspSerArgProLysLeuProGlnThrLysSerArgG 78
    |||||||
162  AAGGACAAAGAGACTCTGACCCCAACTGCCACGACCTCTCCAGAG 211
    |||||||
78  yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrrGluLysAlaLeuTy 95
    |||||||
212  TTGGGCTACCACTCATCTGAGCTGACATATGAAACAGCTCATATTA 261
    |||||||
95  ySserLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspG 111
    |||||||
262  AATCCAAAGACAAGCAAAACCTTGATGATTAATTCATCAGCTTGATG 311
    |||||||
112  CysProHisSerGlnAlaLeuLysValPheAlaGluAsnLysGluI 128
    |||||||
312  TGCCACACAGTCAAGCTTTAAAGAAAGTTTGCTGAATAAAGAAAT 361
    |||||||
128  eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrrGlu 145
    |||||||
362  CCAGAAATGGCAGAGCAGTGTGCTCTCATCTGTTATATGAAACA 411
    |||||||
145  hrAspLysHisLeuSerProAspGlyGlnTyrrValProArgTLeu 161
    |||||||
412  CTGACAAACACCTTCTCTGATGCGCAGTATGTCGCCAGATATGTT 461
    |||||||
162  ValAspProSerLeuThrValArgAlaAspThrArgTyrrSerAs 178
    |||||||
462  GTTGACCATCTCTGACAGTTAGACCGATACACTGGAAGATTAATCAA 511
    |||||||
178  nArgLeuTyrrAlaTyrrGluProAlaAspThrAlaLeuLeuAspAs 195
    |||||||
512  TCGTCTATGCTTACGAACCTGCAGATACAGCTGTGCTTGACAA 561
    |||||||
195  eLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
    |||||||
562  TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATG 596
    |||||||

seq_name: gb_est1:AW582256

seq_documentation_block:
LOCUS AW582256 689 bp mRNA linear EST 16-MAR-2000
DEFINITION OVA-ST0212-120100-075-e10 ST0212 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW582256
VERSION AW582256.1 GI:7257305
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 689)
TITLE HGCP http://www.ludwig.org.br/ORESTES.
JOURNAL The FAPESP/LICR Human Cancer Genome Project
COMMENT Unpublished (1999)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QVA&st=QVA-ST0212-
120100-075-e10&st3=2000-01-12&st4=1)

```

```

Seq primer: puc 18 forward
High quality sequence stop: 687.
FEATURES
    source
        1..689
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="ST0212"
            /dev_stage="Adult"
            /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
            Site_2: SmaI; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters Patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            profiles into the puc 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."
BASE COUNT      212 a      172 c      143 g      162 t
ORIGIN
alignment_scores:
    Quality: 195.00      Length: 195
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-674-266a-181 x AW582256
Align seg 1/1 to: AW582256 from: 1 to: 689
12  GTPProHisProSerArgArgLeuThrGlnGlyArgTrrValArgLysSe 28
    |||||||
25  GGACCGCATCTAGCGCCGACCTCACACAGGACAGTGGTGAGGAAATC 74
    |||||||
28  rargValaIaMetGluLysIleProValSerAlaPheLeuLeuValA 45
    |||||||
75  CAGAGTCCCATGGAGAAATTCAGTGTACGATCTTGCTCTGG 124
    |||||||
45  lalSerTyrrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
    |||||||
125  CCCCTCTCTACACTCTGCGCCAGATACACAGTCAACCTGGAGCCAA 174
    |||||||
62  LysAspThrLysAspSerArgProLysLeuProGlnThrLysSerArgG 78
    |||||||
175  AAGGACAAAGAGACTCTGACCCCAACTGCCACGACCTCTCCAGAG 224
    |||||||
78  yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrrGluLysAlaLeuTy 95
    |||||||
225  TTGGGCTACCACTCATCTGAGCTGACATATGAAACAGCTCATATTA 274
    |||||||
95  ySserLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspG 111
    |||||||
275  AATCCAAAGACAAGCAAAACCTTGATGATTAATTCATCAGCTTGATG 324
    |||||||
112  CysProHisSerGlnAlaLeuLysValPheAlaGluAsnLysGluI 128
    |||||||
325  TGCCACACAGTCAAGCTTTAAAGAAAGTTTGCTGAATAAAGAAAT 374
    |||||||
128  eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrrGlu 145
    |||||||
375  CCAGAAATGGCAGAGCAGTTGTCTCTCATATGTTATGAAACA 424
    |||||||
425  CTGACAAACACCTTCTCTGATGCGCAGTATGTCGCCAGATATGTT 474
    |||||||
162  ValAspProSerLeuThrValArgAlaAspThrArgTyrrSerAs 178
    |||||||
475  GTTGACCATCTCTGACAGTTAGACCGATATCACTGGAAGATTAATCAA 524
    |||||||
178  nArgLeuTyrrAlaTyrrGluProAlaAspThrAlaLeuLeuAspAs 195
    |||||||
525  TCGTCTATGCTTACGAACCTGCAGATACAGCTGTGCTTGACAA 574
    |||||||

```

195 etlyslsAlaLeuLysLeuLysThrGluLeu 206
 |||
 575 TGAAGAAAGCTCTCAAGTTCTGAGACTGAATTG 609

seq_name: gb_est2:B1523582

seq_documentation_block:

LOCUS B1523582 624 bp mRNA linear EST 29-AUG-2001
 DEFINITION 603051712F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201331 5',
 mRNA sequence.

ACCESSION B1523582 GI:15348374

VERSION B1523582.1

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 624)
 NIH-MGC http://mhc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

COMMENT Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM11504 row: h column: 04
 High quality sequence start: 5
 High quality sequence stop: 624.

FEATURES
 source

1..624

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5201331"

/clone_lib="NIH_MGC_122"

/lab_host="DH10B"

/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
 Site-1: NotI; Site-2: EcoRV (destroyed); RNA source
 anonymous pool of 24 week female lung, 16 week female
 spleen, and 20-22 week male spleens. Library is oligo-dT
 primed and directionally cloned (EcoRV site is destroyed
 upon cloning). Average insert size 1.4 kb, insert size
 range 1-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research genetics tracking code 026. Note:
 this is a NIH_MGC Library."

BASE COUNT 190 a 161 c 126 g 147 t
 ORIGIN

alignment_scores:

Quality: 194.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

Length: 194

Gaps: 0

US-09-674-266a-181 x B1523582

Align seg 1/1 to: B1523582 from: 1 to: 624

13 ProHisProSerArgLeuThrGlnGlyArgTrpValAlaGlySerAr 29
 |||
 2 CCGATCTGAGCCGCGACGACGACGAGCGGCGTGAAGAAATCGAG 51
 |||
 29 gvalAlaMeGluLysIleProValSerAlaPheLeuLeuValAla 46
 |||
 52 AGTTGCATGAGAAATTCAGTGTGAGCATTTCTCTGTTGCGCC 101
 |||
 46 euserTyThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62

|||||
 102 TCTCCTACACTGTGGCCAGATACCACTCAAACTGGAGCCAAAG 151
 |||
 63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgLyr 79
 |||
 152 GACACAAAGGACTCTGCAGCCAAACTGCCAGACCTCTCGAGAGTTG 201
 |||

79 pglYAspGlnLeuIleThrPheGlnThrTyrgLugLAlaLeuTyrllys 96
 |||
 202 GGGTGACCAACTCATCTGACTGACATATGAAGAAGCTCTATATAAT 251
 |||

96 erLysThrSerAnLysProLeuMetIleIleHisLysLeuAspGluCys 112
 |||
 252 CCAAGACACCAACCAACCTTGATGATTTATTCATCATCTTGATGATGC 301
 |||

113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 129
 |||
 302 CCACACAGTCACACTTAAAGAAAGTGTGGCTGAATAATGAACAAATCCA 351
 |||

129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgLutThrA 146
 |||
 352 GAATTTGGCAGAGCGATTGTCTCTCAATCTGTTTATGAACAACTG 401
 |||

146 sPLysHisLeuSerProAspGlyGlnTyValProArgIleMetPheVal 162
 |||
 402 ACAACACCTTCTCTCGATGGCCAGATGTCCGCCAGATTATGTTGTT 451
 |||

163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTrpSerAnar 179
 |||
 452 GACCACTCTGACAGTTAGAGCCGATATCATCTGCAAGATATCAATCAATCG 501
 |||

179 gLeuTyAlaTyrgLutProAlaAspThrAlaLeuLeuAsnMetL 196
 |||
 502 TCTTATGCTTACGACCTGACATACAGTCTGTCTGTTGACACAATGCA 551
 |||

196 yslsAlaLeuLysLeuLysThrGluLeu 206
 |||
 552 AGAAGCTCTCAAGTTCTGAGACTGAATTG 583

seq_name: gb_est2:BG179160

seq_documentation_block:

LOCUS BG179160 750 bp mRNA linear EST 06-FEB-2001
 DEFINITION 602330032F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4431583 5',
 mRNA sequence.

ACCESSION BG179160 GI:12685943

VERSION BG179160.1

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 750)
 NIH-MGC http://mhc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

COMMENT Tissue Procurement: DCTD/DRP
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM10186 row: o column: 08
 High quality sequence stop: 732.

FEATURES
 source

1..750

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4431583"

/clone_lib="NIH_MGC_91"

/lissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 234 a 174 c 158 g 183 t 1 others
ORIGIN

alignment_scores:
Quality: 191.00 Length: 191
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266a-181 x BG179160 ..

Align seg 1/1 to: BG179160 from: 1 to: 750

```

15 ProSerArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
|||||
15 CCTAGCCGCGACTCACACAAAGGAGGTGGAGGAAATCCAGAGTTGC 64
|||||
31 aMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerT 48
|||||
65 CATGGAGAAATTCAGTGTCCAGCATTTCTGCTCTGGGCGCTCTCT 114
|||||
48 yTThrLeuAlaArgAspThrValLysProGlyAlaLysLysAspThr 64
|||||
115 ACACCTCTGGCCAGAGATACACAGTCAAACTGGAGCCAAAAGACACA 164
|||||
65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
|||||
165 AAGGACTCTGCAGCCAACTGCCAGACCTCTCCAGAGTTGGGGTGA 214
|||||
81 pGlnLeuIleTrpThrGlnThrTyrgLugLAlaLeuTyLysSerLysT 98
|||||
215 CCAACTCATCTGACTCAGACATATGAAGACCTGATATTAATCCAAAGA 264
|||||
98 hSerAsnLysProLeuMetIleIleHisLysLeuAspLysProHis 114
|||||
265 CAGACAAACACCTTGATGATATTATTCATCTGATGATGAGCCACAC 314
|||||
115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLys 131
|||||
315 AGTCAAAGCTTTAAAGAAAGTGTTCCTGTAATAAAGAAATCCAGAAAT 364
|||||
131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrgLugThrAspLysH 148
|||||
365 GGCAGAGCAGTTGTCCTCTCAATCTGTTATGAAACATGACGACAAAC 414
|||||
148 lLeuSerProAspGlyGlnTyrgValProArgIleMetPheValAspPro 164
|||||
415 ACCTTTCTCTGATGAGCCAGTATGTCGCCAGATATATGTTTGTGCCCA 464
|||||
165 SerLeuThrValArgAlaAspIleThrGlyArgTySerAsnArgLeuTy 181
|||||
465 TCTTCGACGATTAGAGCCGATATCAGTGAAGATATTCAAATCGTCTCTA 514
|||||
181 rAlaTyrgLugProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLys 198
|||||
515 TCCCTTACGAAGCTGCAGATACAGCTCTGTTGCTGACACATGAGAAGA 564
|||||
198 lAlaLeuLysLeuLysThrGlu 205
|||||
565 CTCTCAAGTTGCTGAAGACTGAA 587
|||||

```

seq_name: gb_est2:BE748141

seq_documentation_block:

LOCUS BE748141 689 bp mRNA linear EST 15-SEP-2000

DEFINITION 601571509F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:3838347 5',
mRNA sequence.
ACCESSION BE748141
VERSION BE748141.1 GI:10162133
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC <http://mgc.ncl.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LNCM525 row: a column: 04
High quality sequence stop: 643.
Location/Qualifiers

FEATURES

source
1..689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3838347"
/clone_id="NIH_MGC_55"
/lissue_type="from acute myelogenous leukemia"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
Site: 1: SfiI (ggccgctggcgc); Site: 2: SfiI (ggccatagggc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATATGAGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGGCGCAGCGCGCCGACATG-CT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT 209 a 170 c 148 g 162 t
ORIGIN

alignment_scores:
Quality: 190.00 Length: 190
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266a-181 x BE748141 ..

Align seg 1/1 to: BE748141 from: 1 to: 689

```

15 ProSerArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
|||||
3 CCTAGCCGCGACTCACACAAAGGAGGTGGAGGAAATTCAGAGTTGC 52
|||||
31 aMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerT 48
|||||
65 CATGGAGAAATTCAGTGTCCAGCATTTCTGCTCTGGGCGCTCTCT 102
|||||
48 yTThrLeuAlaArgAspThrValLysProGlyAlaLysLysAspThr 64
|||||
103 ACACCTCTGGCCAGAGATACACAGTCAAACTGGAGCCAAAAGACACA 152
|||||
65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
|||||
153 AAGGACTCTGCAGCCAACTGCCAGACCTCTCCAGAGTTGGGGTGA 202
|||||

```



```

184  uProAlaAspThrAlaLeuLeuAspAsnMetIysAlaLeuLysL 201
|||||
501  ACCGTCAGATACAGCTCTGTTGTCATGACATGAGAAAGCTCTCAAGT 550
201  euleuLysThrGluLeu 206
|||||
551  TGCCTGAAGACCTGAATG 567

seq_name: gb_est1:AW956284

seq_documentation_block:
LOCUS      AW956284              721 bp      mRNA      linear      EST 01-JUN-2000
DEFINITION EST368354 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW956284
VERSION    AW956284.1  GI:8145967
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 721)
            Hegde, P., Ol, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt
            I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeaman, T.J. and
            Quackenbush, J.
            Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
JOURNAL    Unpublished (2000)
COMMENT    The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: John@etlgr.org
            Plate: 90
            Seq primer: Reverse.

FEATURES
            source          1..721
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
                           /clone_lib="MAGE resequences, MAGD"
                           /note="Vector: pBluescriptSkm"
BASE COUNT  224 a      173 c      148 g      175 t      1 others
ORIGIN
alignment_scores:
            Quality: 189.00      Length: 189
            Ratio: 1.000      Gaps: 0
            Percent Similarity: 100.000      Percent Identity: 100.000

Alignment_block:
US-09-674-266a-181 x AW956284 ..
Align seg 1/1 to: AW956284 from: 1 to: 721

18  ArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluLys 34
|||||
21  CGAGTCACACAAAGCAGGTGGGTGAGAAATCCAGATGCCATGAGAGA 70
34  sllEProValSerAlaPheLeuLeuValAlaLeuSerTyrrTrLeuA 51
|||||
71  AATTCAGATGTCAGCATCTTGCTCTGTGGCCCTCTCTACACTCTGG 120
51  laArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSer 67
|||||
121  CCAGAGATACCACTCAAACTGAGCCAAAGACCAAGACCAAGACTCT 170
68  ArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeu1 84
|||||
171  CGACCCAAACTGCCACAGACCTCTCCAGAGGTGGGGTGACCAACTCAT 220
84  eTrpThrGlnThrTyrrGlnGluAlaLeuTyrrLysSerLysThrSerAsnL 101
|||||

```

```

221  CTGACACTCAGACATATGAAAGCTATATTAATCCAAAGCAAGCAACA 270
101  ysProLeuMetIleIleHisIstLeuAspGluCysProHisSerGlnAla 117
|||||
271  AACCTTCATATATATTCATCATCTTGATGATGATGCCACAGACAGCAAGCT 320
118  LeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGlu1 134
|||||
321  TTAAAGAAAGTGTGGCGAAATAAGAAATCCAGAAATGGCAGAGCA 370
134  nPheValLeuLeuAsnLeuValTyrrGluThrThrAspLysIstLeuSerP 151
|||||
371  GTTGTCTCTCTCAATCTGGTTTATGAAACACAGTGCACAAACCTTTCTC 420
151  roAspGlyGlnTyrrValProArgIleMetPheValAspProSerLeuThr 167
|||||
421  CTGATGGCCACTATGTCCCGAGATATATGTTGTGACCCATCTCTGACA 470
168  ValArgAlaAspIleThrGlyArgTyrrSerAsnArgLeuTyrrAlaTyrrG1 184
|||||
471  GTTAGAGCCGATATCATCAGTGAAGATATCAATCGTCTCTATGCTTAGCA 520
184  uProAlaAspThrAlaLeuLeuLeuAspAsnMetIysAlaLeuLysL 201
|||||
521  ACCTGCAGATACAGCTCTGTTGTCATGACATGAGAAAGCTCTCAAGT 570
201  euleuLysThrGluLeu 206
|||||
571  TGCCTGAAGACCTGAATG 587

seq_name: gb_est2:BE796856

seq_documentation_block:
LOCUS      BE796856              948 bp      mRNA      linear      EST 20-SEP-2000
DEFINITION 601588166F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3942112 5',
            mRNA sequence.
ACCESSION  BE796856
VERSION    BE796856.1  GI:10218156
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 948)
            NIH-MGC http://mgc.ncl.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapds-remail.nih.gov
            Tissue Procurement: DCTD/DTF
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
            Plate: LCMW95 row: d column: 17
            High quality sequence stop: 748.

FEATURES
            source          1..948
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
                           /clone_image="3942112"
                           /clone_lib="NIH-MGC-7"
                           /tissue="small cell carcinoma"
                           /cell_line="MGC3"
                           /lab_host="DH10B (phage-resistant)"
                           /note="Organ: Lung; Vector: pOTB7; Site:1: XhoI; Site_2:
                           EcoRI; cDNA made by oligo-dT priming. Directionally
                           cloned into EcoRI/XhoI sites using the following 5'
                           adaptor: GGACGAG(G). Size-selected >500bp for average
                           insert size 1.8kb. Library constructed by Ling Hong in
                           the laboratory of Gerald M. Rubin (University of
                           California, Berkeley) using ZAP-cDNA synthesis kit

```

BASE COUNT 296 a 225 c 207 g 220 t
 ORIGIN (Stratogene) and Superscript II RT (Life Technologies)."

alignment_scores:
 Quality: 187.00 Length: 187
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-266a-181 x BE796856 ..

Align seg 1/1 to: BE796856 from: 1 to: 948

```

20  ThGlglyArgrTpValArGlySerArGValAlaMetGluLysIlePr 36
    |||||||
3  ACACAGGCGAGTGGGTGAGGAAATCCAGAGTGCATGGAAGAAATTC 52
36  oValSerAlaPheLeuLeuValAlaLeuSerTyrThrLeuAlaArg 53
    |||||||
53  AGGTGACGACTTCTGCTGCTGCGCTCTCTACACTGCGCCAGAG 102
53  spThrValLysProGlyAlaLysLysAspThrLysAspSerArGPro 69
    |||||||
103  ATACACAGTCAACCTGGAGCCAAAAGACACAAAAGACTCTCGACCC 152
70  LysLeuProGlnThrLeuSerArGlyTTPGlyAspGlnLeuIleTPrh 86
    |||||||
153  AAACCTGCCAGACCCTCTCCAGAGTTGGGTGACCAACTCATCTGGAC 202
86  rGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnLysProL 103
    |||||||
203  TCAGACATATGAAGAAGCTCTATATTAATCCAGACACAGCAAAACCT 252
103  eumetIleIleHisLysLeuAspGluCysProHisSerGlnAlaLeuLys 119
    |||||||
253  TGATGATTAATTCACCTTGATGAGTGCCACACAGTCAAGCTTTAAAG 302
120  LysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluInPheVa 136
    |||||||
303  AAAGTGTTCCTGAAAATAAAGAAATCCAGAAATGGCAGAGCAGTTGT 352
136  lLeuLeuAsnLeuValTyrGlnThrAspLysHisLysSerProAspG 153
    |||||||
353  CCTCCCAATCTGTTTATGAACAACCTGACAAACACCTTCTCTGATG 402
153  LysIleTyrValProArGileMetPheValAspProSerLeuThrValArg 169
    |||||||
403  GCCAGTATGTCGCCAGGATATGTTGTTCACCATCTCGACAGTTAGA 452
170  AlaAspIleThrGlyArGlyTyrSerAsnArgLeuTyrAlaTyrGluProAl 186
    |||||||
453  GCCGATATCTCTGGAAGATATTCAAACCCCTCTATGCTTACGAACTGC 502
186  aAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysLeuLeuL 203
    |||||||
503  AGATACAGCTCTGCTGCTTGACACATGAAAGAAAGCTCTCAAGTCTGA 552
203  ysrThrGluLeu 206
    |||||||
553  AGACTGAATTG 563
seq_name: gb_est2:BE870718
seq_documentation_block:
LOCUS BE870718 734 bp mRNA linear EST 20-OCT-2000
DEFINITION 60148570F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852448 5',
ACCESSION BE870718
VERSION BE870718.1 GI:10319507
KEYWORDS EST.
SOURCE human.
```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 734)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LLM9574 row: 1 column: 17
High quality sequence stop: 723.
FEATURES
source Location/Qualifiers
1..734
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3852448"
/clone_1lb="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Colon; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 231 a 163 c 156 g 184 t
ORIGIN
alignment_scores:
Quality: 186.00 Length: 186
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-266a-181 x BE870718 ..
Align seg 1/1 to: BE870718 from: 1 to: 734
21  GlnGlyArgrTpValArGlySerArGValAlaMetGluLysIleProVa 37
    |||||||
2  CAAGCAGGTGGGTGAGGAAATCCAGAGTGCATGGAAGAAATTCAGT 51
37  lSerAlaPheLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspT 54
    |||||||
52  GTACAGACTTCTGCTGCTTGCGCCCTCTCTACACTGCGCCAGAGATA 101
54  hrThrValLysProGlyAlaLysLysAspThrLysAspSerArGProLys 70
    |||||||
102  CCACAGTCAAACTGGAGCCAAAAGACACAAAAGACTCTCGACCCAAA 151
71  LeuProGlnThrLeuSerArGlyTTPGlyAspGlnLeuIleTPrhArgI 87
    |||||||
152  CTGCCCCAGACCCTCTCCAGAGTTGGGTGACCAACTCATCTGGACTCA 201
87  nThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnLysProLeuM 104
    |||||||
202  GACATATGAAGAAAGCTCTATATTAATCCAAAGACACAAACCCCTTGA 251
104  etIleIleHisLysLeuAspGluCysProHisSerGlnAlaLeuLysLys 120
    |||||||
252  TGATTAATTCACCTTGAGATGAGTGCACACAGTCAAGCTTTAAAGAAA 301
121  ValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluInPheValLe 137
    |||||||
302  GTGTTTGTGAAAATAAAGAAATCCAGAAATTTGGCAGAGCAGTTTGCTC 351
137  uLeuAsnLeuValTyrGlnThrAspLysHisLysSerProAspLysG 154
```

```
|||||
352 CCTCAATCTGGTTATGAACACGACAAACACCTTCTCCTGATGGCC 401
154 InTyValProArgIleMetPheValAspProSerLeuThrValArgAla 170
402 AGTATGTCCTCCAGGATATGTTTGTGGACCATCTCTCAGATTAGACC 451
171 AspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGluProAlaAs 187
452 GATATCACTGAGAGATATTCAACCGCTCTATGCTTAGAACCTGCAGA 501
187 pThrAlaLeuLeuAspAsnMetLysLysAlaLeuLysLeuLeuLysT 204
502 TACAGCTCTGTGCTGTGACAAATGAAAGAGCTCTCAAGTTGCTGAGA 551
204 hrGluLeu 206
552 CTGAATTG 559

seq_name: gb_est2:BG285246

seq_documentation_block:
LOCUS BG285246 866 bp mRNA linear EST 21-FEB-2001
DEFINITION 602409569F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4539353 5',
mRNA sequence.
ACCESSION BG285246
VERSION BG285246.1 GI:13037011
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 866)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10467 row: 1 column: 18
High quality sequence stop: 750.
FEATURES
Source
1..866
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="NIH_MGC_91"
/tissue_type="adenoecarcinoma, cell line"
/ab_host="DH10B (phage-resistant)"
/Note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 270 a 194 c 189 g 212 t 1 others
ORIGIN
alignment_scores:
Quality: 184.00 Length: 184
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-266A-181 x BG285246 ..
Align seg 1/1 to: BG285246 from: 1 to: 866
```

```
17 ArgArgLeuThrGlnGlyArgTyrValArgLysSerArgValAlaMetG1 33
30 CGCGGACATCACACAAGGAGGTGGGTGAGAAATCCAGATTGCCATGGA 79
33 uLysIleProValSerAlaPheLeuLeuValAlaLeuSerTyrThrL 50
80 GAAATTCATGATGTCAGCATCTTGTGCTCTGTGGCCCTCTCCATACATC 129
50 euAlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAsp 66
130 TGGCCAGATACCAACAGTCAACCTGAGCCAAAAGAGACAAAGAGAC 179
67 SerArgProLysLeuProGlnThrLeuSerArgGlyTyrGlyAspGlnLe 83
180 TCTGCACCCCAACAGTCCCGACAGCCCTCCAGAGAGTGGGTGACCAACT 229
83 uLleTyrThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerA 100
230 CATCTGAGACTCAGACATATGAAGAAGCTCTATATATCAACAGACAGCA 279
100 snLysProLeuMetIleIleHisLysLeuAspGluCysProHisSerGln 116
280 ACAAACTTGTATGATATATTCATCCTTGATGAGTGCACACAGCTCA 329
117 AlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaG1 133
330 GCTTAAAGAAAGTGTGCTGAGAAATTAAGAAATCCAGAAATTGGCAGA 379
133 uGlnPheValLeuLeuAsnLeuValTyrGluThrAspLysHisLys 150
380 GCAGTTTCTCTCCCAATCTGTTATGAACAACATCAACACCCCTTT 429
150 erProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeu 166
430 CTCTGATGGCCAGTATGTCCCGAGATTATGTTGTGACCATCTGTG 479
167 ThrValAlaGalaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyr 183
480 ACAGTTAGAGCGGATATCATCGAAGATATTCAAATGCTCTATGCTTA 529
183 rGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuL 200
530 CGAACCTGACATACAGCTCTGTGCTTGACACACATGAAGAAAGCTCTCA 579
200 ys 200
580 AG 581

seq_name: gb_est2:BG386151

seq_documentation_block:
LOCUS BG386151 751 bp mRNA linear EST 12-MAR-2001
DEFINITION 602455333F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:458336 5',
mRNA sequence.
ACCESSION BG386151
VERSION BG386151.1 GI:13279597
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 751)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
```

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 plate: LCM1307 row: b column: 09
 High quality sequence stop: 672.
 Location/Qualifiers

FEATURES

1. 751

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4583356"
 /clone_1lb="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 230 a 181 c 153 g 187 t
 ORIGIN

alignment_scores:
 Quality: 178.00 Length: 178
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-266a-181 x BG386151 ..

Align seg 1/1 to: BG386151 from: 1 to: 751

```

29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
|||||
26 AGAGTTGCATGAGAAATTCAGTGTCCAGCATTTGCTCCTTGTCG 75
|||||
45 aleuSetTyThrLeuAlaArgAspThrThrValLysProGlyAlaLys 62
|||||
76 CCTCTCTACACTCTGGCCAGAGTACACAGTCAAACTGGAGCCAAA 125
|||||
62 ysaSPThrLysAspSerArgProLysLeuProGlnThrLysSerArgGly 78
|||||
126 AGGACAAAGAGACTCTGCACCAACTGCCCGACCCCTTCCAGAGCT 175
|||||
79 TrpGlyaspGlnLeuIleTrpThrGlnThrTygGluGluAlaLeuTy 95
|||||
176 TGGGTTGACCAACTCATCTGAGCTCAGACATATGAGAGCTTATATA 225
|||||
95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuaspGlu 112
|||||
226 ATCCAAAGACAAGCAAAACCTTGATGATATTCACATCAGTATGAT 275
|||||
112 ySPProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGlu 128
|||||
276 GCCACACAGTCAAGCTTTAAAGAAAGTGTGCTGAATAAAAGAAATC 325
|||||
129 GlnLysLeuAlaGluGlnPheValLeuAsnLeuValTyrgLutThr 145
|||||
326 CAGAAATTTGGAGAGAGAGTTGTCTCTCAATGTGGTTATGAACAAC 375
|||||
145 rAspLysHisLeuSerProaspGlyGlnTyValProArgIleMetPhe 162
|||||
376 TGACAAACACCTTTCTCTGATGGCCAGATATGCCAGAGTATGTGG 425
|||||
162 alaSPProSerLeuThrValArgAlaAspIleThrGlyArgTySerAsn 178
|||||
426 TTGACCATCTCTGACAGTTAGACCGATATCAGTGGAGATATTAAC 475
|||||
179 ArgLeuTyAlaTyrgLutProAlaAspThrAlaLeuLeuAspAsnMe 195
|||||
476 CGTCTATATGCTTAGGAACCTGCAGATACACGCTGTGTTGCTTGACA 525
|||||

```

195 tLysLysAlaLeuLysLeuLeuTyThrGluLeu 206
 |||||
 526 GAAGAAGACTCTCAAGTTGCTGGAAGACTGAATG 559

seq_name: gb_est2:BM006356

seq_documentation_block:

LOCUS BM006356 758 bp mRNA linear EST 30-OCT-2001
 DEFINITION 603615055F1 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5421109 5',
 mRNA sequence.

ACCESSION BM006356
 VERSION BM006356.1 GI:16520710
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 758)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

plate: LCM1876 row: m column: 14
 High quality sequence stop: 754.

Location/Qualifiers

FEATURES

1. 758

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5421109"
 /clone_1lb="NIH_MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 238 a 176 c 156 g 188 t
 ORIGIN

alignment_scores:
 Quality: 178.00 Length: 178
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-266a-181 x BM006356 ..

Align seg 1/1 to: BM006356 from: 1 to: 758

```

29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
|||||
31 AGAGTTGCATGAGAAATTCAGTGTCCAGCATTTGCTCCTTGTCG 80
|||||
45 aleuSetTyThrLeuAlaArgAspThrThrValLysProGlyAlaLys 62
|||||
81 CCTCTCTACACTCTGGCCAGAGTACACAGTCAAACTGGAGCCAAA 130
|||||
62 ysaSPThrLysAspSerArgProLysLeuProGlnThrLysSerArgGly 78
|||||

```

```

131 AGGACGACAAAGGACCTCGACCCCAACCTGCCACAGCCCTCTCCACAGGT 180
79 TTPGlyAspGlnLeuIleTrpThrGlnThrTYRGlUGlnAlaLeuTyrly 95
181 TGGGGTGACCAACTCATCTGAGCTCGACATATGAAGAAGCTCTATATATA 230
95 sSerLystrSerAsnLySProleuNetIleIleHsHsLeuAspGluC 112
231 ATCCAAAGACACAGCAACAAACCTTGATGATTTATTCATCACTGGATGAGT 280
112 yAfBrHsSerGlnAlaLeuLysLysValpheaIaGluAsnLySgIuIle 128
281 GCCCACACAGTCACAGCTTTAAAGAAAGTGTTTCTGAATAATAAGAAATC 330
129 GlnLyLeuAlaGlnGlnInPheValIleLeuLeuAsnLeuValTYRclunThrH 145
331 CAGAAATTTGGCAGAGCATTTTGCTCCCTCAATCTGGTTATGAAACACAC 380
145 rAspLyHsHsLeuSerProAspLyGlnTyrValIProAqIleMetPheV 162
381 TGACAAACACCTTTCTCCGATGCCAGATATGCCACAGATATATGTTTG 430
162 aLAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
431 TTGACCCATCTCTGCACAGTTAGACCGCAATATCACTCGAAGATATTCAAAC 480
179 ArgLeuTYRAlATyrGlnProAlaAspTrpAlaLeuLeuLeuSpasne 195
481 CGCTCTATGCTTACCAACCTGCAGATACAGCTCTTGTCTTGCACACAT 530
195 cLysLySAlaLeuLysLeuLeuLysThrGluLeu 206
531 GAAGAAAGCTCTCAAGTGTCTGAAAGCTGAATG 564

seq_name: gb_est2:BM006510

seq_documentation_block:
LOCUS BM006510 797 bp mRNA linear EST 30-OCT-2001
DEFINITION 603615461F1 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5421399 5',
            mRNA sequence.
ACCESSION BM006510
VERSION BM006510.1 GI:16520864
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 797)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMN at:
http://image.llnl.gov
plate: L16M1877 row: 1 column: 16
High quality sequence stop: 792.
location/Qualifiers
1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5421399"
/clone_id="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directly cloned into EcoRI/XhoI sites using the

```

following 5' adaptor: GGACAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

```
alignment_scores:
  Quality: 178.00      Length: 178
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.0000
```

```
alignment_block:
US-09-674-266A-181 x BM006510
```

Align seg 1/1 to: BM006510 from: 1 to: 797

	29	ArgValAlaMetGlnLysIleProValSerAlaPheLeuLeuValAl	45
	31	AGAGTTGGCATGTGAGAAATTCACGTGCTAGCACTTCTCCTTGTCG	80
	45	aLeuSerTyThrLeuAlaArgAspThrThrValLysProGlyAlaLysL	62
	81	CCTCTCTACTACTCTGTGCCAGAGATTCCACAGCCAACCTGGAGCCAAA	130
	62	yAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly	78
	131	AGGACACAAGGACTCTGCACCMAAATCGCCAGACCTCTCCAGAAGGT	180
	79	TTPGLYASGLNLEULETRPHRGINTTHRYGLUGLNALALEUYYRLY	95
	181	TGGGGTGACCACTCATCTGGACTCGACATAAGAAAGCTCTATATTA	230
	95	sSerLysThrSerAsnLysProLeuMetIleIleHisLysLeuAPGUC	112
	231	ATCCAAAGAACACCAAAACCTTGATGATTATTCATCACCTTGGATGACT	280
	112	ysFroHISerGlnAlaLeuLysValPheAlaGlnAsnLysGluLe	128
	281	GCCCACACAGTCAGCTTTAAAGAAAGTGTTTGCTGAATAAAGAAATC	330
	129	GlnLysLeuAlaGlnGlnPheValLeuLeuAsnLeuValTYrGluThrRh	145
	331	CAGAAATTTGGCAGAGCAGATTGTCCCTCCCAATCTGGTTATGAACAAC	380
	145	rAspLysHisLeuSerProAspGlyGlnTyrrValProArgIleMetPheY	162
	381	TGCACAAACACCTTTCTCTGATGGCCAGATGTCCCCAGAGATTATTTTG	430
	162	aAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn	178
	431	TTCAGCCATCTCTGACAGTTAGAGCCGATATCATCTGGAAGATTTCAAAC	480
	179	ArgLeuTyrrAlaTyrrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe	195
	481	CGTCTCTATGCTTAGCAACCTGCAGATPAACGCTCTGCTTGACAAACAT	530
	195	tLysLysAlaLeuLysLeuLeuLysThrGlnLeu	206
	531	GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTTG	564
seq_name:	gb_est2:BM006784		
seq_documentation_block:			
LOCUS	BM006784	802 bp	mRNA linear EST 30-OCT-2001
DEFINITION	603615217.F1 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5420836 5'		
	mRNA sequence.		
ACCESSION	BM006784		
VERSION	BM006784.1 GI:16521138		
KEYWORDS	EST.		
SOURCE	human.		

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 802)
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L1CM1876 row: b column: 05
 High quality sequence stop: 795.
 Location/Qualifiers
 1..802
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5420836"
 /clone_lib="NIH-MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."
 BASE COUNT 257 a 180 c 162 g 203 t
 ORIGIN
 alignment_scores: Quality: 178.00 Length: 178
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-674-266a-181 x BM006784 ..
 Align seg 1/1 to: BM006784 from: 1 to: 802

331 CAGAAATGGCAGACAGTTGTTCCTCCATCTGGTTATATGAACAC 380
 145 TASPVSHisLeuSerProAspGlyGlnTyValProAlaIleMetPheV 162
 381 TGACAAACACCTTTCCTGATGGCCAGTATGCCAGGATATATGTCG 430
 162 aLAspProSerLeuThrValArgAlaAspIlePheGlyArgTySerAsn 178
 431 TTGACCATCTCTGACAGTTAGAGCCGATATACCTGGAAGATATTCAA 480
 179 ArgLeuTyArgIatArgIupProAlaAspThrAlaLeuLeuAspAsn 195
 481 CGTCTATCTCTTACGACACCTGCAGATACAGCTCTGTGTTGACAA 530
 195 TlYSLySLAlaLeuLysLeuLeuLysThrGluLeu 206
 531 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 564
 seq_name: gb_est2:BM006554
 seq_documentation_block:
 LOCUS BM006554 846 bp mRNA linear EST 30-OCT-2001
 DEFINITION 603615521P1 NIH-MGC_110 Homo sapiens cDNA clone IMAGE:5421413 5', mRNA sequence.
 ACCESSION BM006554
 VERSION BM006554.1 GI:16520908
 KEYWORDS EST.
 SOURCE
 ORGANISM Homo sapiens
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 846)
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L1CM1877 row: j column: 06
 High quality sequence stop: 781.
 Location/Qualifiers
 1..846
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5421413"
 /clone_lib="NIH-MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."
 BASE COUNT 278 a 192 c 173 g 203 t
 ORIGIN
 alignment_scores: Quality: 178.00 Length: 178
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:

US-09-674-266a-181 x BM006554 ..

Align seg 1/1 to: BM006554 from: 1 to: 846

```
29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
|||||
31 AGAGTTGGCATGAGAAAATTCAGATGACATTCCTGCTGTGGC 80
45 aleuSerThrLeuAlaArgAspThrThrValLysProGluAlaLysL 62
|||||
81 CCTCTCCACACTCTGGCCAGATACACAGCTCAAACTGGAGCCAAA 130
62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
|||||
131 AGGACACAAAGAGCTCTGACCAAACTGCCCAAGCCCTCTCCAGAGCT 180
79 TrpGlyAspGlnLeuIleTyrThrGlnThrTyrGlnGluAlaLeuTyrLy 95
|||||
181 TGGGGTGACCACTCATCTGGACTGACATATGAGAGAGCTTATATTA 230
95 sSerLysThrSerAnLysProLeuMetIleIleHisLysLeuAspGluC 112
|||||
231 ATCCAGACAGCAAGCAACACCCTTGATGATATTCATCCTTGATGAGT 280
112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
|||||
281 GCCCAGACAGCTCAAGCTTTAAAGAAAGTGTTCGAAATAAAGAAATC 330
129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
|||||
331 CAGAAATTGGCAGAGCAGTTGTCTCCTCAATCTGGTTTATGAAACAC 380
145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
|||||
381 TGACAAACACCTTCTCCTGATGGCCAGATGTCCCAAGATTATGTTTG 430
162 alaSPProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
|||||
431 TTGACCCCATCTCTGACAGTTAGAGCCGATATCATCGAAGATATTCAAC 480
179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuAspAsnMe 195
|||||
481 CGTCTCTATGCTTACGAACTGAGATACAGCTGTGTGCTTGACACACT 530
195 tLysLysAlaLeuLysLeuLysThrGluLeu 206
|||||
531 GAAGAAGGCTCTCAAGTTGCTGAGAGCTGAATTG 564
```


OM of: US-09-674-266A-181 to: Issued_Patents_NA: * out_format : pfs

Date: Oct 8, 2002 11:10 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framer.p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09674266/runat.08102002_093409_27017/app_query.fasta.1.271
-DB=Issued_Patents_NA -QPM=fasta -SUFFIX=olip2n.rni
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPEL=0.000
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=60.000
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000
-YGAPEXT=60.000 -DELDP=6.000 -DELEXT=7.000 -START=1
-MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
-THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09674266.@CGN1.1.46 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-674-266A-181
Query length: 206
Database: Issued Patents_NA: *
Database sequences: 38353
Database length: 122816752
Search time (sec): 46.220000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation	
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-1		192.00	3613.53	1.4e-193	875	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-247-155-61		178.00	3343.21	1.6e-178	1889	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-35		107.00	2009.44	3.1e-104	386	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-040-984-78		102.00	1909.16	1.2e-98	733	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-123-912-78		102.00	1909.16	1.2e-98	733	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-221-298-32		66.00	1232.37	5.9e-61	940	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-247-155-149		62.00	1150.01	2.3e-56	940	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-36		53.00	986.63	2.9e-47	373	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-607-209		52.00	974.26	1.4e-46	159	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-439-313-209		52.00	974.26	1.4e-46	159	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-37		51.00	953.36	1.6e-45	158	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-9		38.00	700.35	2.5e-31	489	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-10		37.00	681.14	3.0e-30	506	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-38		33.00	615.49	1.4e-26	136	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-5		24.00	432.00	2.2e-16	732	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-11		14.00	248.71	3.6e-06	329	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-12		14.00	248.71	3.6e-06	329	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-13		14.00	248.71	3.6e-06	329	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-14		14.00	248.71	3.6e-06	329	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-15		14.00	248.68	3.7e-06	330	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-42		11.00	207.38	0.0007	44	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-44		11.00	206.55	0.0008	49	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-41		11.00	192.30	0.0030	311	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-41		10.00	173.38	0.0572	310	1
/cgn2_6/ptodata1/1na/6A.COMB.seq:US-09-023-082A-41		7.00	135.08	7.77	28	1
/cgn2_6/ptodata1/1na/6A.COMB.seq:US-09-023-082A-4		7.00	133.36	9.68	35	1
/cgn2_6/ptodata1/1na/6A.COMB.seq:US-09-023-082A-6		7.00	133.36	9.68	35	1
/cgn2_6/ptodata1/1na/6A.COMB.seq:US-09-023-082A-12		7.00	117.30	75.99	281	1
/cgn2_6/ptodata1/1na/5B.COMB.seq:US-08-757-653-121		7.00	117.30	75.99	281	1
/cgn2_6/ptodata1/1na/5B.COMB.seq:US-08-757-653-127		7.00	116.20	87.48	324	1
/cgn2_6/ptodata1/1na/5B.COMB.seq:US-09-171-461-51		7.00	115.90	90.95	337	1
/cgn2_6/ptodata1/1na/5B.COMB.seq:US-08-756-386-56		7.00	115.90	90.95	337	1
/cgn2_6/ptodata1/1na/5B.COMB.seq:US-08-823-516-45		7.00	115.90	90.95	337	1
/cgn2_6/ptodata1/1na/6A.COMB.seq:US-08-682-853A-56		7.00	115.90	90.95	337	1
/cgn2_6/ptodata1/1na/6A.COMB.seq:US-08-759-038-56		7.00	115.90	90.95	337	1
/cgn2_6/ptodata1/1na/6A.COMB.seq:US-08-758-314-56		7.00	115.90	90.95	337	1
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-350-309-56		7.00	115.90	90.95	337	1
/cgn2_6/ptodata1/1na/5B.COMB.seq:US-08-691-814B-113		7.00	115.48	96.02	356	1

/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-020-956-160	7.00	114.97	102.42	380
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-030-607-160	7.00	114.97	102.42	380
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-09-439-313-160	7.00	114.97	102.42	380
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-794-002-5	7.00	112.35	143.38	534
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-854-039B-5	7.00	112.35	143.38	534
/cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-854-039B-1	7.00	111.50	159.84	596
/cgn2_6/ptodata1/1na/5A.COMB.seq:US-08-406-248-3	7.00	111.49	160.10	597

seq_name: /cgn2_6/ptodata1/1na/6B.COMB.seq:US-08-916-576B-1

seq_documentation_block:

Sequence 1, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: TIO, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916.576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 875 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 71..595
NAME/KEY: mat_peptide
LOCATION: 131..595
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 71..130
US-08-916-576B-1

alignment_scores:

Quality:	192.00	Length:	192
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-674-266A-181 x US-08-916-576B-1

Align seg 1/1 to: US-08-916-576B-1 from: 1 to: 875

```

15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
   |||
20 CCGAGCCCGCAGCTCACACAGGAGGTGGGAGGAAATCCAGAGTTC 69
   |||
31 ametGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerT 48
   |||
70 CATGAGAGAAATTCAGTGCAGCATCTCTGCTGCGCCCTCTCCT 119
   |||
48 yThrLeuAlaArgAspThrValLysProGlyAlaLysLysAspThr 64
   |||
120 ACACCTGTGGCAGATACACAGTCAACCTGGAGCAAAAAAGACACA 169
   |||
65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
   |||
170 AAGGACTCTCCAGCCCAAACTGCCAGACCTCTCCAGAGTGGGGTGA 219
   |||
81 pGluLeuIleTrpThrGlnThrTyrgluGluAlaLeuTyrlsSerLysT 98
   |||
220 CCAACTCATGTGACTCAGACATATGAGAAGCTCTATATTAATCCAGA 269
   |||
98 hrSerAsnLysProLeuMetIleIleHisLysLeuAspGluCysProHis 114
   |||
270 CAAGCAACAAACCCCTTGATGATTAATTCATCATCTGGATGAGCCAC 319
   |||
115 SerGlnAlaLeuLysValAlaPheAlaGluAsnLysGluIleGlnLysLe 131
   |||
320 AGTCAGCTTTAAAGAAAGTGTCTCTAATAATAAGAAATCCAGAAAT 369
   |||
131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrgluThrThrAspLysH 148
   |||
370 GGCAGAGAGGTTGTCTCTCAATCTGTTATGAAACAACTGACAAAC 419
   |||
148 lSLeuSerProAspGlyLysIleTyrgluProArgIleMetPheValAspPro 164
   |||
420 ACCTTTCTCCGATGGCCAGATATGCCAGGATATATGTTGTTACCCA 469
   |||
165 SerLeuThrValArgAlaAspIleThrGlyArgTyrlsSerAsnArgLeuTy 181
   |||
470 TCTCGACAGATTAGACCGCATATCTGGAAGATTTCAAAATCGTCTCTA 519
   |||
181 rAlaTyrgluProAlaAspThrAlaLeuLeuAsnMetLysLysLys 198
   |||
520 TGCTTACCAACCTGCAGATACAGCTCTGTGCTGTGCAACAATGAGAAG 569
   |||
198 lAlaLeuLysLeuLeuLysThrGluLeu 206
   |||
570 CTCTCAAGTTGCTGAAGACTGAATGG 595

```

seq_name: /cgn2_6/plodata/1/lna/5B_COMB.seq:US-09-247-155-61

seq_documentation_block:

```

: Sequence 61, Application US/09247155A
: Patent No. 6312922
: GENERAL INFORMATION:
: APPLICANT: Dumas Milne Edwards, Jean-Baptiste
: APPLICANT: Duclert, Aymeric
: APPLICANT: Bougueret, Lydie
: TITLE OF INVENTION: Complementary DNAs
: FILE REFERENCE: GENSET.021A
: CURRENT APPLICATION NUMBER: US/09/247.155A
: CURRENT FILING DATE: 1999-02-09
: EARLIER APPLICATION NUMBER: 60/074.121
: EARLIER FILING DATE: 1998-02-09
: EARLIER APPLICATION NUMBER: 60/081.563
: EARLIER FILING DATE: 1998-04-13
: EARLIER APPLICATION NUMBER: 60/096.116
: EARLIER FILING DATE: 1998-08-10
: EARLIER APPLICATION NUMBER: 60/099.273
: EARLIER FILING DATE: 1998-10-04
: NUMBER OF SEQ ID NOS: 182
: SOFTWARE: Patent.pm

```

```

: SEQ ID NO 61
: LENGTH: 1689
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 51..575
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 51..110
: OTHER INFORMATION: Von Heijne matrix
: OTHER INFORMATION: score 11.2
: OTHER INFORMATION: seq AFLTVALSYTLA/RD
: FEATURE:
: NAME/KEY: polyA_signal
: LOCATION: 1653..1658
: FEATURE:
: NAME/KEY: polyA_site
: LOCATION: 1674..1689
: US-09-247-155-61

```

```

alignment_scores:
  Quality: 178.00      Length: 178
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment_block:

US-09-674-266a-181 x US-09-247-155-61 ..

Align seg 1/1 to: US-09-247-155-61 from: 1 to: 1689

```

29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
   |||
42 AGAGTTCAGATGAGAGAAATTCAGATGTCAGCATTTCTGCTGTGGC 91
   |||
45 aleuSerTyrlsLeuAlaArgAspThrValLysProGlyAlaLysL 62
   |||
92 CCTCTCTACACTGTGGCAGATACACAGTCAAACTGGAGCCAAAA 141
   |||
62 yAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
   |||
142 AGGACACAAGAGACTCTGACCCAAACTGCCACAGCCCTCCAGAGGT 191
   |||
79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrgluAlaLeuTyrls 95
   |||
192 TGGGTGACCAACTCATCTGCACTCAGACATATGAGAAGCTCTATATA 241
   |||
95 sSerLysThrSerAsnLysProLeuMetIleIleHisLysLeuAspGluC 112
   |||
242 ATCCAAGACAAGCAACAACCTTGATGATATATCATCACTTGATGAGT 291
   |||
112 ySProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluLe 128
   |||
292 GCCCACACAGTCAACCTTTAAAGAAAGTGTGCGAAATAAAGAAATC 341
   |||
129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrgluThr 145
   |||
342 CAGAAATTTGGCAGACAGTGTCTCTCTCAATCTGTTTATGAAGAAC 391
   |||
145 rAspLysHisLeuSerProAspGlyLysIleTyrgluProArgIleMetPhe 162
   |||
392 TGACAAACACCTTTCTCGATGGCCAGATAGTCCCAAGATATAGTTTG 441
   |||
442 TTGACCACTCTCTGACAGTTAGCCGATATCACGGAAGATATCAAAAT 491
   |||
162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrlsSer 178
   |||
179 ArgLeuTyrlsAlaTyrgluProAlaAspThrAlaLeuLeuAspAsnMe 195
   |||
492 CGTCTCTATGCTTACGAACTGCGAGATACAGCTCTGTTGCTTGACAA 541
   |||
195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206

```



```
30 1AlametglulysileProvalSerAlaPheleuleuValAlaLeus 47
   |||||||
52 TGCCATGGAGAAATTCAGTGTGACATCTTGTCTCTGTGGCCCTCT 101
   |||||||
47 eTyrThrleuAlaArgAspThrThrValLysProGlyAlaLysLysasp 63
   |||||||
102 CCTACACTCTGGCCAGAGATACACAGTCMAACCTGAGCCAAAAAGAGC 151
   |||||||
64 ThrLysAspSerArgProLysleuProGlnThrleuSerArgGlyTrpG1 80
   |||||||
152 ACAAAAGACTCTGACCAACCACTGCCAGACCTCTCCAGAGGTTGGGG 201
   |||||||
80 yAspGlnleuIleTrrPThrGlnThrTyrgluGluAlaLeuTyrlsSerL 97
   |||||||
202 TGACCACTCATCTGACTCGACATGACATATGAGAAGAACTCATATTAATCA 251
   |||||||
97 ystHSerAsnLysProleuMetIleIleHsHisleuAspGluCysPro 113
   |||||||
252 AGACAAGCAACMAACCTTGATGATTATTCATCACTTGATGATGAGGCCA 301
   |||||||
114 HsSer 115
   |||||||
302 CACAGT 307

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-123-912-78

seq_documentation_block:
; Sequence 78, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C1
; CURRENT APPLICATION NUMBER: US/09/123.912A
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/040.802
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 78
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (309)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (492)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (563)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (657)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (660)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (703)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (708)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (710)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (711)
; OTHER INFORMATION: Where n is a, c, g or t
```

```
; NAME/KEY: modified_base
; LOCATION: (732)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (740)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (748)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (758)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (762)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (765)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (787)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-123-912-78
```

```
alignment_scores:
  quality: 102.00      Length: 102
  Ratio: 1.000         Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000
```

```
alignment_block:
US-09-674-266A-181 x US-09-123-912-78 ..
```

```
Align seg 1/1 to: US-09-123-912-78 from: 1 to: 793
```

```
14 HsPProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
   |||||||
2 CATCTAGCCGCCGACTCACACMAAGCAGTGGAGAGAAATTCAGAGT 51
   |||||||
30 1AlametglulysileProvalSerAlaPheleuleuValAlaLeus 47
   |||||||
52 TGCCATGGAGAAATTCAGTGTGACATCTTGTCTCTGTGGCCCTCT 101
   |||||||
47 eTyrThrleuAlaArgAspThrThrValLysProGlyAlaLysLysasp 63
   |||||||
102 CCTACACTCTGGCCAGAGATACACAGTCMAACCTGAGCCAAAAAGAGC 151
   |||||||
64 ThrLysAspSerArgProLysleuProGlnThrleuSerArgGlyTrpG1 80
   |||||||
152 ACAAAAGACTCTGACCAACCACTGCCAGACCTCTCCAGAGGTTGGGG 201
   |||||||
80 yAspGlnleuIleTrrPThrGlnThrTyrgluGluAlaLeuTyrlsSerL 97
   |||||||
202 TGACCACTCATCTGACTCGACATGACATATGAGAAGAACTCATATTAATCA 251
   |||||||
97 ystHSerAsnLysProleuMetIleIleHsHisleuAspGluCysPro 113
   |||||||
252 AGACAAGCAACMAACCTTGATGATTATTCATCACTTGATGATGAGGCCA 301
   |||||||
114 HsSer 115
   |||||||
302 CACAGT 307

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-221-298-32

seq_documentation_block:
; Sequence 32, Application US/09221298
; Patent No. 6284241
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471
; CURRENT APPLICATION NUMBER: US/09/221.298
```

```
: CURRENT FILING DATE: 1998-12-23
: NUMBER OF SEQ ID NOS: 112
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 32
: LENGTH: 401
: TYPE: DNA
: ORGANISM: Human
US-09-221-298-32

alignment_scores:
  Quality: 66.00      Length: 80
  Ratio: 0.835        Gaps: 1
  Percent Similarity: 98.750  Percent Identity: 98.750

alignment_block:
US-09-674-266a-181 x US-09-221-298-32  ..

Align seg 1/1 to: US-09-221-298-32 from: 1 to: 401

111 GluCySPROHISerGlnAlaLeuLysValPheAlaGlnAsnLysG1 127
|||||
164 GAGTGGCCACACAGTCAGCTTAAAGAAAGTGTGCTGAAATVAAAGA 213
|||||
127 vllEglnLysLeuAlaGlnPheValLeuLeuAsnLeuValTYrG1UT 144
|||||
214 AATCCGAATATTGGCAGACAGTTTCTCTCCCAATCTGGTTTATGAAA 263
|||||
144 hTThraspLysLysSerProAspGlyGlnTYrValProArGileMet 160
|||||
264 CAAGTCAGAACACCTTCTCTCGATGGCCAGATATGCCAGAGATTANG 313
|||||
161 PheValAspProSerLeuThrValArgAla.AspliethrGlyArTYrs 177
|||||
314 TTTGTTGACCCATCTCTGACAGTTAGAGCCCGATATACACTGGAAGATATT 363
|||||
177 eFAsnArgLeuTYrAlaTYrGlnProAlaAspThrAla 189
|||||
364 CAACCGTCTCTATGCTTACGAACCTGCAGATACAGCT 401

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-247-155-149

seq_documentation_block:
: Sequence 149, Application US/09247155A
: Patent No. 6312822
: GENERAL INFORMATION:
: APPLICANT: Dumas Milne Edwards, Jean-Baptiste
: APPLICANT: Duclet, Aymeric
: APPLICANT: Bouquellet, Lydie
: TITLE OF INVENTION: Complementary DNAs
: FILE REFERENCE: GENSET.021A
: NUMBER OF APPLICATIONS: US/09/247.155A
: CURRENT FILING DATE: 1999-02-09
: EARLIER APPLICATION NUMBER: 60/074.121
: EARLIER FILING DATE: 1998-02-09
: EARLIER APPLICATION NUMBER: 60/081.563
: EARLIER FILING DATE: 1998-04-13
: EARLIER APPLICATION NUMBER: 60/096.116
: EARLIER FILING DATE: 1998-08-10
: EARLIER APPLICATION NUMBER: 60/099.273
: EARLIER FILING DATE: 1998-10-04
: NUMBER OF SEQ ID NOS: 182
: SOFTWARE: Patent.pm
: SEQ ID NO 149
: LENGTH: 940
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 177..569
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 177..236
```

```
: OTHER INFORMATION: Von Hejme matrix
: OTHER INFORMATION: score 11.199998092651
: OTHER INFORMATION: seq AFLIVLSTYLA/RD
: FEATURE:
: NAME/KEY: poly_site
: LOCATION: 931..939
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 482
: OTHER INFORMATION: n=a, g, c or t
US-09-247-155-149

alignment_scores:
  Quality: 62.00      Length: 62
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-674-266a-181 x US-09-247-155-149  ..

Align seg 1/1 to: US-09-247-155-149 from: 1 to: 940

29 ArgValAlaMetGlnLysIleProValSerAlaPheLeuLeuValAl 45
|||||
168 AGAGTGGCCATGGAAATAATTCAGTCAGCATCTTCTCTCTGTGGC 217
|||||
45 aLeuSerTYrThrLeuAlaArgAspThrValLysProGlyAlaLysL 62
|||||
218 CCTCTCTACACTCTGGCCAGATACACAGTCACAACTGGAGCCANAA 267
|||||
62 yAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgLy 78
|||||
268 AGGACACAAAGAGACTCTGACCCAAACTGCCAGACCTCTCCAGAGGT 317
|||||
79 TrpGlyAspGlnLeuIleTrrPThrGlnTrhTYrGlu 90
|||||
318 TGGGTGACCAACTCATCTGAGACACARACATATGAA 353

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-36

seq_documentation_block:
: Sequence 36, Application US/08916576B
: Patent No. 6171816
: GENERAL INFORMATION:
: APPLICANT: YU, GUO-LIANG
: APPLICANT: DILLON, PATRICK J.
: APPLICANT: EBNER, REINHARD
: APPLICANT: ENDRESS, GREGORY A.
: TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
: STREET: 1100 NEW YORK AVENUE, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: US
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/916.576B
: FILING DATE:
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/024.347
: FILING DATE: 23-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: STEFFE, ERIC K.
: REGISTRATION NUMBER: 36.688
```

```
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 373 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-916-576B-36

alignment_scores:
Quality: 53.00 Length: 53
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x US-08-916-576B-36 ..
Align seg 1/1 to: US-08-916-576B-36 from: 1 to: 373

16 SerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaIle 32
1 AGCCGCCGACTCACAGAGGCGAGTGGTGAGGAATCCAGAGTTGCCAT 50
32 tGtLysIleProValSerAlaPheLeuLeuValAlaLeuSerTyrT 49
51 GGAGAAATTCAGAGTCAGCATCTCTGCTGCGCCCTCTCTCA 100
49 hrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThrLys 65
101 CTGCGCCGAGATACACAGCTCAACCTGAGCCAAAAGACACAAAG 150
66 AspSerArg 68
151 GACTCTCGA 159

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-030-607-209

seq_documentation_block:
; Sequence 209, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
```

```
INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-030-607-209

alignment_scores:
Quality: 52.00 Length: 52
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x US-09-030-607-209 ..
Align seg 1/1 to: US-09-030-607-209 from: 1 to: 159

25 ValArgLysSerArgValAlaMetGluLysIleProValSerAlaPheLe 41
3 GTGAGGAAATCCAGAGTTGCCATGAGAAATTCAGAGTCAGCATTTCTT 52
41 uLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspThrThrValLysP 58
53 GCTCTGTTGGCCCTCTCTCACTGCTGCGCAGAGATACACAGTCAAC 102
58 rGlyAlaLysLysAspThrLysAspSerArgProLysLeuProGlnThr 74
103 CTGGAGCCAAAAGACACAAAGACTCTGACCCAACTGCCCCAGACC 152
75 LeuSer 76
153 CTCTCC 158

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-439-313-209

seq_documentation_block:
; Sequence 209, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kaios, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 209
; LENGTH: 159
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-439-313-209

alignment_scores:
Quality: 52.00 Length: 52
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
```

US-09-674-266a-181 x US-09-439-313-209 ..

Align seg 1/1 to: US-09-439-313-209 from: 1 to: 159

```
25 ValArgLysSerArgValAlaMetGluLysIleProValSerAlaPheLeu 41
|||||
3 GTGAGGAAATCCAGAGTTGCCATGAGAAATTCAGTGCAGCATTTCTT 52
|||||
41 uleuLeuValAlaLeuSerTyrThrIleuAlaArgAspThrThrValLysP 58
|||||
53 GCTCCTGTGGCCCTCTCTACACTCTGGCCAGATACACACAGTCAAAAC 102
|||||
58 roGlyAlaLysAspThrLysAspSerArgProLysLeuProGlnThr 74
|||||
103 CTGAGCCCAAAAGACACAAAGACTCTGACCCCAAACTGCCCCAGACC 152
|||||
75 LeuSer 76
|||||
153 CTCCTCC 158
```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-37

seq_documentation_block:

```
; Sequence 37, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-916-576B-37
```

alignment_scores:
Quality: 51.00 Length: 51
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266a-181 x US-08-916-576B-37 ..

Align seg 1/1 to: US-08-916-576B-37 from: 1 to: 158

```
15 ProSerArgArgLeuThrGlnGlyArgTyrValArgLysSerArgValAla 31
|||||
6 CCTAGCCCGCGACTCACACAAAGCAGGTGGGTGAGGAATTCACAGTTGC 55
|||||
31 aMetGluLysIleProValSerAlaPheLeuLeuValAlaLeuSerT 48
|||||
56 CATGAGAAATTCAGTGCAGCATTTGCTCTGTGGCCCTCTCTCT 105
|||||
48 yThrIleuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
|||||
106 ACACCTGTGGCCAGAGATACACAGTCAAACTGGAGCCAAAGACACACA 155
|||||
65 Lys 65
|||||
156 AAG 158
```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-9

seq_documentation_block:

```
; Sequence 9, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-916-576B-9
```

alignment_scores:
Quality: 38.00 Length: 38
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-266A-181 x US-08-916-576B-9 ..

Align seg 1/1 to: US-08-916-576B-9 from: 1 to: 489

```
29 ArgvalAlaMetGluLysIleProValSerAlaPheLeuLeuValAl 45
|||||
40 ACAGTTGCATGAGAAATTCACAGTCACGATCTCTCTGTGAC 89
|||||
45 AleuSerYrThrLeuAlaArgAspThrValLysProGlyAlaLysL 62
|||||
90 CCTCTCTACACTGTGGCCAGAGATACACAGTCAAACTGGAGCCAAA 139
62 yAspThrLysAsp 66
|||||
140 AGGACACAAAGAC 153
```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-10

seq_documentation_block:

```
; Sequence 10, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 506 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-916-576B-10
```

alignment_scores:

```
Quality: 37.00 Length: 37
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:

US-09-674-266A-181 x US-08-916-576B-10 ..

Align seg 1/1 to: US-08-916-576B-10 from: 1 to: 506

```
109 LeuAspGluCysProHisSerGlnAlaLeuLysLysValPheAlaGlu 125
|||||
50 TTGGATGAGTGCACACAGCTTAAAGAAAGTGTGGTGAANA 99
|||||
125 nLysGluIleGlnLysIleuAlaGluGlnPheValLeuLeuAsnLeuValT 142
|||||
100 TAAAGAAATCCAGAAATTTGGCAGAGAGATTGTCTCTCATCTGGTTT 149
142 yrGluThrThr 145
|||||
150 ATGAACAACACT 160
```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-38

seq_documentation_block:

```
; Sequence 38, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-916-576B-38
```

alignment_scores:

```
Quality: 33.00 Length: 33
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:

US-09-674-266A-181 x US-08-916-576B-38 ..

Align seg 1/1 to: US-08-916-576B-38 from: 1 to: 136

13 ProHisProSerArgLeuThrGlnGlyArgTyrValArgLysSer 29
 2 CGCCATCTTAGCCGCGACTACACAAAGCAGTGGGTGAGAAATCCAG 51
 29 GValAlaMetGluLysIleProValSerAlaPheLeuLeuValAla 45
 52 AGTTGGCATGAGAAATTCAGTGCAGCATCTTCTCTGTGGCC 100
 seq_name: /cgn2.6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-5

seq_documentation_block:
 : Sequence 5, Application US/08916576B
 : Patent No. 6171816
 : GENERAL INFORMATION:
 : APPLICANT: YU, GUO-LIANG
 : APPLICANT: DILLON, PATRICK J.
 : APPLICANT: EBNER, REINHARD
 : APPLICANT: ENDRESS, GREGORY A.
 : TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
 : NUMBER OF SEQUENCES: 45
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P. L. L. C.
 : STREET: 1100 NEW YORK AVENUE, SUITE 600
 : CITY: WASHINGTON
 : STATE: DC
 : COUNTRY: US
 : ZIP: 20005-3934
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/916, 576B
 : FILING DATE:
 : CLASSIFICATION: 536
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 60/024,347
 : FILING DATE: 23-AUG-1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: STEFFE, ERIC K.
 : REGISTRATION NUMBER: 36,688
 : REFERENCE/DOCKET NUMBER: 1488.0500001
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202) 371-2600
 : TELEFAX: (202) 371-2540
 : INFORMATION FOR SEQ ID NO: 5:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 732 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: double
 : TOPOLOGY: linear
 : MOLECULE TYPE: CDNA
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: 49..546
 : FEATURE:
 : NAME/KEY: mat_peptide
 : LOCATION: 118..546
 : FEATURE:
 : NAME/KEY: sig_peptide
 : LOCATION: 49..117
 : US-08-916-576B-5

alignment_scores:
 Quality: 24.00 Length: 24
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-266a-181 x US-08-916-576B-5 ..

Align seg 1/1 to: US-08-916-576B-5 from: 1 to: 732
 149 LeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 165
 373 TTATCACTGATGGCGCAATATGTGCTAGAAATCATGTTGTAGACCTTC 422
 165 rLeuThrValArgAlaAspIle 172
 423 TTTAACAGTTAGAGCTGACATA 444

